

**STIC-Biotech/ChemLib**

---

**From:** Duffy, Patricia  
**Sent:** Tuesday, June 19, 2001 12:23 PM  
**To:** STIC-Biotech/ChemLib  
**Subject:** Sequence search 09/438,185  
**Importance:** High

In re: 09/438,185

Please search SEQ ID NO:1047.  
Please backtranslate SEQ ID NO:1047 into NA and run against NA databases.  
Please perform an interference search and print out top 25 hits.

Thank you,  
**Patricia A. Duffy**  
**Art Unit 1645**  
**Office CM-1, 8D05**  
**Mailbox CM-1, 8E12**  
**Phone 703-305-7555**

**Point of Contact:**  
**Mona Smith**  
**Technical Info. Specialist**  
**CM1 12C14 Tel: 308-3278**

Point of Contact:  
Mona Smith  
Technical Info. Specialist  
CMI 12014 Tel: 308-3278

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2001, 11:40:08 ; Search time 62.02 Seconds  
(without alignments)  
353.852 million cell updates/sec

Title: US-09-438-185-1047

Perfect score: 1889

Sequence: 1 VHCERTLPDKILKILK.....ESIPLYNQEKYLSGFEVLCQ 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Sum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_0601.\*

1:	/SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT.*
2:	/SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT.*
3:	/SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT.*
4:	/SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT.*
5:	/SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT.*
6:	/SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT.*
7:	/SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT.*
8:	/SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT.*
9:	/SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT.*
10:	/SIDSL/gcgdata/geneseq/geneseq/AA1989.DAT.*
11:	/SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT.*
12:	/SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT.*
13:	/SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT.*
14:	/SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT.*
15:	/SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT.*
16:	/SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT.*
17:	/SIDSL/gcgdata/geneseq/geneseq/AA1996.DAT.*
18:	/SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT.*
19:	/SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT.*
20:	/SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT.*
21:	/SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT.*
22:	/SIDSL/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1250.5	66.2	259	20	AA135703
2	222	11.8	498	15	AA150185
3	221	11.7	498	14	AA136741
4	220	11.6	497	15	AA150184
5	219	11.6	497	14	AA136740
6	213.5	11.3	452	12	AA131119
7	213.5	11.3	452	18	AA131119
8	213.5	11.3	452	20	AA131119
9	213.5	11.3	452	21	AA131119
10	206.5	10.9	532	21	AA131327
11	99	5.2	924	19	AA169848

12	98.5	5.2	1024	18	AA19604
13	97	5.1	439	20	AA135368
14	95	5.0	920	21	AA131679
15	95	5.0	944	21	AA131678
16	95	5.0	970	21	AA131677
17	94	5.0	293	18	AA155671
18	94	5.0	1226	18	AA131325
19	92	4.9	3923	20	AA131327
20	92	4.9	4536	20	AA141262
21	92	4.9	4536	20	AA156826
22	91.5	4.8	872	18	AA126605
23	91.5	4.8	872	20	AA131812
24	91.5	4.8	872	20	AA131813
25	91.5	4.8	872	20	AA131816
26	91	4.8	2183	19	AA148707
27	91	4.8	2183	19	AA148708
28	91	4.8	2183	19	AA148709
29	91	4.8	2183	19	AA148710
30	91	4.8	2183	19	AA148710
31	90.5	4.8	615	16	AA172276
32	90.5	4.8	615	18	AA122228
33	90	4.8	892	20	AA102528
34	90	4.8	2183	14	AA139592
35	89	4.7	392	19	AA153839
36	89	4.7	392	21	AA152603
37	88	4.7	438	18	AA140373
38	88	4.7	438	18	AA101572
39	88	4.7	438	21	AA184036
40	88	4.7	659	21	AA184047
41	88	4.7	720	21	AA184049
42	88	4.7	749	21	AA184048
43	87.5	4.6	309	20	AA139785
44	87.5	4.6	328	20	AA139789
45	87.5	4.6	521	20	AA139789

#### ALIGNMENTS

RESULT 1  
AA135703  
ID AA135703 standard; Protein; 259 AA.  
XX  
AC AA135703;  
XX  
DT 13-SEP-1999 (first entry)  
XX  
DE Chlamydia pneumoniae transmembrane protein sequence.  
XX  
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;  
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;  
KW vaccine; neutralising epitope.  
XX  
OS Chlamydia pneumoniae.  
XX  
PN WO927105-A2.  
XX  
PD 03-JUN-1999.  
XX  
PF 20-NOV-1998; 98WO-IB01890.  
XX  
PR 04-NOV-1998; 98US-0107078.  
XX  
PR 21-NOV-1997; 97FR-0014673.  
XX  
PA (GEST ) GENSET.  
XX  
PI Griffiths R;  
XX  
DR WPI; 1999-357842/30.  
XX  
PT Genome sequence of Chlamydia pneumoniae  
XX  
PS Page 1410-1411; Disclosure; 1912pp; English.

XX AAV34584-Y35879 represent the proteins encoded by all the open reading  
 CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.  
 CC C. pneumoniae causes respiratory disease such as pneumonia and  
 CC bronchitis and is thought to be a contributing factor in heart  
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema  
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading  
 CC frames of the C. pneumoniae genome (see AAV34584-Y35879) can be used in  
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae  
 CC nucleotide sequences can also be used as immunogenic compositions,  
 CC especially where the vector directs the expression of a neutralising  
 CC epitope of C. pneumoniae.  
 XX Sequence 259 AA;

Query Match 66.2%; Score 1250.5; DB 20; Length 259;  
 Best Local Similarity 91.8%; Pred. No. 3.7e-118;  
 Matches 245; Conservative 3; Mismatches 8; Indels 11; Gaps 1;

QY 96 VVSTPFNNRWYLLSSRLSKSCYCPFFIDYLEAFGLLSDFLDHQAIVKPFLETHF 155  
 DB 4 mvstpfllvsmeklls-----kifdyleafgllsdfldhgvikfelethf 52  
 QY 156 SYYPVSGFVAPHOYLSLQDRYFPPIASVMRTLDKNFSLTPDLIHDLGHVPWLLHPSFS 215  
 DB 53 SYYPVSGFVAPHGYLSLQDRYFPPIASVMRTLDKNFSLTPDLIHDLGHVPWLLHPSFS 112  
 QY 216 EFFINMGRLFTKVIKVOALPSKKORIOTLOSNIATVRCFWFTVESGLIENHGRKAYG 275  
 DB 113 effinmgrlftkviakvqalpskkqriqltqslniaivrcfwftvesgllienhgrkayg 172  
 QY 276 AVLISSPQELGHAFIDNVRVLPLELDQIIRLPENTSTPOETLFSRHFDLVELTSKLEW 335  
 DB 173 avliisspqlghafidnvrvlpleldqirpntstpoetlfsrhfdelveltsklew 232  
 QY 336 MDQGLLESIPLYNOEKYLSGFEVLQ 362  
 DB 233 mldqgllesiplynqekylsgfevlq 259

RESULT 2  
 AAR50185  
 ID AAR50185 standard; Protein; 498 AA.

XX AAR50185;  
 DE 04-NOV-1994 (first entry)

XX Rat tyrosine hydroxylase.

XX Human; rat; tyrosine hydroxylase; TH; substitution; N-terminal;  
 KW enhanced enzymatic activity; wild-type; transformation; N-terminal;  
 KW Parkinsons disease; Alzheimers disease; brain; encapsulation;  
 KW selectively permeable; polymer capsule; antibody.

XX Rattus rattus.

XX US5300436-A.

XX 05-APR-1994.

XX 13-MAR-1991; 91US-06569446.

XX 13-MAR-1991; 91US-06569446.

XX 26-JAN-1993; 93US-0009075.

XX (UUNY ) UNIV NEW YORK STATE.

XX Filer D, Friedhoff AJ, Goldstein M, Wu J;

XX WPI; 1994-125849/15.

PT Modified DNA encoding variant tyrosine hydroxylase with an  
 PT N-terminal amino acid substn. and cells contg. it - can be used  
 PT to treat diseases associated with defective function of the  
 PT enzyme, e.g. Parkinson's disease or Alzheimer's disease  
 XX Disclosure; Flig 5; 27pp; English.

XX This sequence represents rat tyrosine hydroxylase (TH) and may be  
 CC used in the production of the variant TH molecules of the invention.  
 CC These variants contain amino acid substitutions in the N-terminal  
 CC portion of the molecule, and in the rat molecule these substitutions  
 CC are pref. at positions Ser8, Ser19, Ser31, Arg38, Glu40, Glu43 or  
 CC Arg46. Variants containing one or more of these amino acid  
 CC substitutions, have substantially enhanced enzymatic activity  
 CC compared to the wild-type enzyme. The most pref. substitution being  
 CC Ser40 for Tyr or Leu. DNA encoding the TH variants, and cells  
 CC transformed with this DNA may be used for treating diseases associated  
 CC with defective function of TH, or dopamine, eg. Parkinsons disease  
 CC and Alzheimers disease, and effective disorders. The cells can then  
 CC be implanted into the brain or encapsulated in a selectively permeable  
 CC polymer capsule which allows release of the cells products but protects  
 CC them from attack by the hosts antibodies or cells.

XX Sequence 498 AA;

Query Match 11.8%; Score 222; DB 15; Length 498;

Best Local Similarity 26.7%; Pred. No. 7.1e-14;

Matches 65; Conservative 39; Mismatches 103; Indels 36; Gaps 5;

QY 112 SSRFSLWKS-----YCPREFLDYLEAFGLLSDFLDH-----QAVIKFFELETHF 155

DB 227 aeelatwkevyvtlkglyathacrehlegfqlerycgyredsipqledvsrfktertfg 286

QY 156 SYYPVSGFVAPHOYLSLQDRYFPPIASVMRTLDKNFSLTPDLIHDLGHVPWLLHPSFS 215

DB 287 qlprvagllsakdflaslafvqctgyirhasspmhspepcchellghvpmldartfa 346

QY 216 EFFINMGRLFTKVIKVOALPSKKORIOTLOSNIATVRCFWFTVESGLIENHGRKAYG 275

DB 347 qfsqdg-----laslgasdeeklst-----vywftvefglckqngelkayg 390

QY 276 AVLISSPQELGHAFIDNVRVLPLELDQIIRLPENTSTPOETLFSRHFDLVELTS 331

DB 391 agllssygelhlsiseepvrafdpdaavpdydqtyqpyfvsvsfndakklrnyas 450

QY 332 KLE 334

DB 451 riq 453

RESULT 3  
 AAR36741  
 ID AAR36741 standard; protein; 498 AA.

XX AAR36741;

XX 20-SEP-1993 (first entry)

XX Rat tyrosine hydroxylase.

XX Human; rat; tyrosine hydroxylase; substitution; N-terminal; dopamine;  
 KW neurological disorder; antibody; variant.

XX Rattus rattus.

XX Key Location/Qualifiers

XX Modified-site 8 /note= "Phosphorylation site"

XX Modified-site 19 /note= "Phosphorylation site"

XX Modified-site 40 /note= "Phosphorylation site"



FT Modified-site 153 /note= "Phosphorylation site"  
 FT US5212082-A.  
 PN 18-MAY-1993.  
 PD 13-MAR-1991; 91US-0669446.  
 XX 13-MAR-1991; 91US-0669446.  
 XX (UUNY ) UNIV NEW YORK STATE.  
 PA Filer D, Friedhoff AJ, Goldstein M, Wu J;  
 XX WPI; 1993-175456/21.  
 XX Genetically modified tyrosine hydroxylase having increased  
 PT activity - used for treating neurological disorders e.g.  
 PT Parkinson's and Alzheimer's diseases and affective disorders  
 XX Disclosure; Fig 5; 20pp; English.

CC The sequences given in AR36740-41 represent the human and rat  
 CC tyrosine hydroxylase protein respectively. These sequences may be  
 CC used in the construction of a variant tyrosine hydroxylase which  
 CC contains at least one amino acid substitution in the N-terminal 55  
 CC amino acids. The substitution corresponds to an amino acid selected  
 CC from Ser8, Ser31, Arg37, Arg38, Gln39, Ser40, Ile42, Glu43,  
 CC Asp44, Ala45, Arg46 or Lys47. The enzymatic activity of the variant  
 CC protein is at least 3-fold greater than that for the wild type  
 CC protein. Cells transfected with the DNA encoding these proteins may  
 CC be used for treating neurological disorders associated with a  
 CC deficiency in tyrosine hydroxylase or dopamine. These proteins may  
 CC be used to generate antibodies specific for the variant tyrosine  
 CC hydroxylases to monitor the enzyme during a treatment regimen.

XX Sequence 498 AA;  
 SQ

Query Match 11.7%; Score 221; DB 14; Length 498;  
 Best Local Similarity 27.4%; Pred. No. 9e-14; Indels 36; Gaps 5;  
 Matches 65; Conservative 37; Mismatches 99;

QY 118 WKS-----YCPREFLDYLEAFGLLSDFLDH-----QAVIKFFLETHFSYYPVS 161  
 DB 233 wkevvytlkgllyathacrehlegfqllycgyredsipqledvsrflktertqrlrpa 292

QY 162 GFVAPHQYLSLQDRYFPPIASVMFTLDKDNFSLTPDLIHDLLGHVPWLLHPSFEPINM 221  
 DB 293 glisakdfilaslafvfgctqyirhasspmhspepcchellghvpmldrtafqfsqdi 352

QY 222 GRLETKVIEKVOALPSKKORTQTLQSNLIAIVRCFWFTVESGLIENHGRKAYGAVLISS 281  
 DB 353 g-----laslgasdeeelek1st-----vywftvefglckngelkayagaglls 396

QY 282 POELGHAFIDNVRVLPLELDQIIRLPNTSPQETLFSIRHF-----DELVELTSKLE 334  
 DB 397 ygellhslepevratpdaaavqpyqddtqpyvfysesfndakdkirnyasriq 453

RESULT 4  
 ID AAR50184 standard; Protein; 497 AA.  
 XX AAR50184;  
 AC AAR50184;  
 DT 04-NOV-1994 (first entry)  
 XX Human tyrosine hydroxylase.  
 XX Human; rat; tyrosine hydroxylase; TH; substitution; N-terminal;  
 KW enhanced enzymatic activity; wild-type; transformation; dopamine;

KW Parkinsons disease; Alzheimers disease; brain; encapsulation;  
 KW selectively permeable; polymer capsule; antibody.  
 XX Homo sapiens.  
 XX US5300436-A.  
 XX 05-APR-1994.  
 XX 13-MAR-1991; 91US-0669446.  
 XX 13-MAR-1991; 91US-0669446.  
 PR 26-JAN-1993; 93US-0009075.  
 XX (UUNY ) UNIV NEW YORK STATE.  
 XX Filer D, Friedhoff AJ, Goldstein M, Wu J;  
 XX WPI; 1994-125849/15.  
 DR N-PSDB; AA044899.  
 XX Modified DNA encoding variant tyrosine hydroxylase with an  
 PT N-terminal amino acid substn. and cells contg. it - can be used  
 PT to treat diseases associated with defective function of the  
 PT enzyme, e.g. Parkinson's disease or Alzheimer's disease  
 XX Disclosure; Column 21-24; 27pp; English.

XX This sequence represents human tyrosine hydroxylase (TH) and may be  
 CC used in the production of the variant TH molecules of the invention.  
 CC These variants contain amino acid substitutions in the N-terminal  
 CC portion of the molecule, and in the human molecule these  
 CC substitutions are pref. at positions Ser19, Ser31, Arg37, Arg38,  
 CC Ser40, Leu41, Glu43, Asp44, Ala45, Arg46 or Lys47. Variants  
 CC containing one or more of these amino acid substitutions, have  
 CC substantially enhanced enzymatic activity compared to the wild-type  
 CC enzyme. The most pref. substitution being Ser40 for Tyr or Leu.  
 CC DNA encoding the TH variants, and cells transformed with this DNA  
 CC may be used for treating diseases associated with defective function  
 CC of TH, or dopamine, eg. Parkinson's disease and Alzheimers disease,  
 CC and effective disorders. The cells can then be implanted into the  
 CC brain or encapsulated in a selectively permeable polymer capsule  
 CC which allows release of the cells products but protects them from  
 CC attack by the hosts antibodies or cells.

XX Sequence 497 AA;  
 SQ

Query Match 11.6%; Score 220; DB 15; Length 497;  
 Best Local Similarity 26.7%; Pred. No. 1.1e-13;  
 Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;

QY 112 SSRSFWKS-----YCPREFLDYLEAFGLLSDFLDH-----QAVIKFFLETHF 155  
 DB 226 aesiactkveytlkgllyathacrehlegfqllycgyredsipqledvsrflktertqf 285

QY 156 SYYPVSGFVAPHQYLSLQDRYFPPIASVMFTLDKDNFSLTPDLIHDLLGHVPWLLHPSFS 215  
 DB 286 qlrpvagllsardfiaslafvfgctqyirhasspmhspepcchellghvpmldrtaf 345

QY 216 EFTINMGRLETKVIEKVOALPSKKORTQTLQSNLIAIVRCFWFTVESGLIENHGRKAYG 275  
 DB 346 qfsqdig-----laslgasdeeelek1stis-----wftvefglckngelkayg 389

QY 276 AVLISSPQELGHAFIDNVRVLPLELDQIIRLPNTSPQETLFSIRHF-----DELVELTS 331  
 DB 390 agllssygeellhclsepeirafpdaaavqpyqddtqpyvfysesfndakdklrsyas 449

QY 332 KLE 334  
 DB 450 riq 452



CC (CT-A- cholera toxin, DT-B'- truncated diphtheria toxin,  
 CC SITA- Shiga-like toxin A; HIVP-BP- HIV protease binding protein.  
 CC See also AAQ12710-12.  
 XX  
 SQ Sequence 452 AA;

Query Match 11.3%; Score 213.5; DB 12; Length 452;  
 Best Local Similarity 26.6%; Pred. No. 4.5e-13;  
 Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;  
 QY 104 RNLWRLSSRLSKSYCPREFLDYLEAFGLLSDFLDH-----QAVIKFFFELETHF 155  
 Db 184 kktwgtvfklsklyktha---cyeynhiflplekycgfhednlpqledvsgflqtctgf 240  
 QY 156 SYYPVSGFVAPHQYLSLQDRYFPFIASVMRTLDKNFSLTPDLIHDLGHVPWLLHPSPFS 215  
 Db 241 rlrpvagllssrdflgglafvfhctqyirhgskpmytpepdichellghvplfsdrsa 300  
 QY 216 EFTINMGRLEFTRVIEKVQALPSKKQRIQTQSNLIAIVRCFWFTVESGLIENHEGRKAYG 275  
 301 qfsqelg-----laslgapdeyieklat-----iywftvefglckggsikayg 344  
 QY 276 AVLISSPOELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHDELVE 328  
 Db 345 agllssfgelqycisekplilelektaiqnyvtvtefqlpyyvaesfndake 397

## RESULT 7

AAW25788  
 ID AAW25788 standard; Protein; 452 AA.

AC AAW25788;  
 XX  
 DT 27-MAR-1998 (first entry)  
 DE Human phenylalanine hydroxylase.  
 KW Phenylalanine hydroxylase; human; hybrid protein; cell delivery;  
 KW cell binding ligand; translocation domain; diphtheria toxin B';  
 KW phenylketonuria; therapy.  
 XX Homo sapiens.  
 OS US5668255-A.  
 PN 16-SEP-1997.  
 PD 07-JUN-1984; 84US-0618199.  
 PF 27-JUN-1991; 91US-0722484.  
 PR 07-JUN-1984; 84US-0618199.  
 PR 25-APR-1985; 85US-0726808.  
 PR 07-JUN-1985; 85US-0742554.  
 PR 22-DEC-1989; 89US-0456095.  
 PR 14-JUN-1990; 90US-0538276.  
 PR 04-AUG-1993; 93US-0102387.  
 XX (SERA-) SERAGEN INC.  
 XX  
 XX Murphy JR;  
 XX  
 XX WPI; 1997-470103/43.  
 XX N-PSDB; AAT91639.  
 XX  
 XX New hybrid molecules for delivery of agents to cells - comprise a  
 PT binding domain of a cell binding ligand and a portion of a  
 PT translocation domain of a protein  
 XX Example 5; Fig 13A-C; 30pp; English.  
 PS  
 XX This protein comprises for human phenylalanine hydroxylase (PH).  
 CC A DNA molecule (see AAT91638) encoding PH was used to construct a

CC PH-diphtheria toxin B' gene that was expressed in E. coli. The  
 CC resulting hybrid protein can be used in the treatment of  
 CC phenylketonuria. The active PH enzyme is targeted to, and  
 CC incorporated into, the broad range of cells which native diphtheria  
 CC normally attacks, achieving the widespread therapy that is needed  
 CC for this inherited disorder. Claimed hybrid proteins comprise a  
 CC translocation domain and a cell binding domain. They can be used  
 CC for the delivery of agents (e.g. therapeutic genes, toxins,  
 CC detectable labels) into cells. The use of a translocation  
 CC mechanism ensures that the hybrid will be effective in relatively  
 CC low doses, since a high proportion of the substance of interest  
 CC will be taken into the targeted cells. The hybrid molecules can be  
 CC manufactured as a single hybrid recombinant protein, permitting  
 CC reproducibility, consistency, and the precise control of  
 CC composition.  
 XX  
 SQ Sequence 452 AA;

Query Match 11.3%; Score 213.5; DB 18; Length 452;  
 Best Local Similarity 26.6%; Pred. No. 4.5e-13;  
 Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps 4;

QY 104 RNLWRLSSRLSKSYCPREFLDYLEAFGLLSDFLDH-----QAVIKFFFELETHF 155  
 Db 184 kktwgtvfklsklyktha---cyeynhiflplekycgfhednlpqledvsgflqtctgf 240  
 QY 156 SYYPVSGFVAPHQYLSLQDRYFPFIASVMRTLDKNFSLTPDLIHDLGHVPWLLHPSPFS 215  
 Db 241 rlrpvagllssrdflgglafvfhctqyirhgskpmytpepdichellghvplfsdrsa 300  
 QY 216 EFTINMGRLEFTRVIEKVQALPSKKQRIQTQSNLIAIVRCFWFTVESGLIENHEGRKAYG 275  
 Db 301 qfsqelg-----laslgapdeyieklat-----iywftvefglckggsikayg 344  
 QY 276 AVLISSPOELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHDELVE 328  
 Db 345 agllssfgelqycisekplilelektaiqnyvtvtefqlpyyvaesfndake 397

## RESULT 8

AAV55893  
 ID AAV55893 standard; Protein; 452 AA.

AC AAV55893;  
 XX  
 DT 15-FEB-2000 (first entry)  
 DE Human phenylalanine hydroxylase.  
 KW Recombinant; hybrid; binding domain; ligand; animal cell; diphtheria;  
 KW translocation domain; botulinum; neurotoxin; ricin; cholera; tetanus;  
 KW shiga-like toxin; pertussis; translocation; cytoplasmic membrane; HIV;  
 KW cytosol; therapy; genetic deficiency disease; enzyme; co-factor; poison;  
 XX adipocyte; cancer; virus; infection; antibody.  
 OS Homo sapiens.  
 XX  
 XX US5965406-A.  
 XX 12-OCT-1999.  
 XX 07-JUN-1995; 95US-0488246.  
 XX 04-AUG-1993; 93US-0102387.  
 PR 07-JUN-1984; 84US-0618199.  
 PR 27-JUN-1991; 91US-0722484.  
 PR 25-APR-1985; 85US-0726808.  
 PR 07-JUN-1985; 85US-0742554.  
 PR 22-DEC-1989; 89US-0456095.  
 PR 14-JUN-1990; 90US-0538276.  
 XX (SERA-) SERAGEN INC.

```

XX      Murphy JR.;
PI      WPI: 1999-632431/54.
DR      N-PSDB; AAZ30664.
PT      Recombinant DNA molecule encoding a three part hybrid protein used in
PT      the treatment of Aids and genetic deficiency diseases -
XX      Example 5; Fig 13; 31pp; English.
XX      The invention relates to a recombinant DNA molecule encoding a hybrid
CC      protein comprising three parts: (a) the first part comprises a portion
CC      of the binding domain of a cell-binding polypeptide ligand allowing the
CC      hybrid protein to bind to an animal cell; (b) the second part comprises
CC      a portion of a translocation domain of a naturally occurring protein
CC      selected from diphtheria toxin, botulinum neurotoxin, ricin, cholera
CC      toxin, LT toxin, C3 toxin, shiga-like toxin, pertussis toxin and tetanus
CC      toxin, which translocate the third part of the across the cytoplasmic
CC      membrane into the cytosol of the cell; and (c) the third part comprises
CC      a polypeptide entity to be introduced into the cell, which is non-native
CC      to the naturally occurring protein of (b). This sequence represents the
CC      human phenylalanine hydroxylase for use in generating the hybrid of the
CC      invention. The hybrid molecule enables the direction of appropriate
CC      therapy to affected cells, allowing them to function properly and
CC      alleviate or cure the disease. The hybrid is especially used in treating
CC      genetic deficiency diseases, by delivering to affected cells an enzyme
CC      supplying the missing function, to supplementing cellular levels of a
CC      particular enzyme or a scarce precursor or cofactor, to directing toxins
CC      or other poisons to destroy particular cells (such as adipocytes, cancer
CC      cell, or virus infected-cells), to counteracting viral infections such as
CC      HIV, by introducing appropriate antibodies to viral proteins. It is also
CC      involved in the process of getting non-therapeutic substances such as
CC      detectable labels into cells.
XX      Sequence 452 AA;
SQ
      Query Match 11.3%; Score 213.5; DB 20; Length 452;
      Best Local Similarity 26.6%; Pred. No. 4.5e-13;
      Matches 62; Conservative 44; Mismatches 100; Indels 27; Gaps
      QY 104 RNLAWRLLSRFSRWKSCPRFLDYLEAGFLSDSLDHH-----QAVIKFFELETHF 155
      DB 184 kktgwrvfktslyktha---cyeynhifplekyqcfhednipledvsgfqlccgcf 240
      CC 156 SYYPVSGFVAPHQVLSLLQDRFPPTASVWRTLDKDNLSLPDLHLLGHVPMWLLHPSPS 215
      DB 241 rlrpvagllsrdoflgslafvrhctqyrhsgkmpytpedpichellghvlfdsrsta 300
      QY 216 EFFINMGRLEFKVIEKVOALPSKKORIQTLOSLNLAIVRCFWFTVESGLIENHGKATG 275
      DB 301 qfsqelg-----laslgapdeyieklat-----iywtfvfeigkqgdsikayg 344
      QY 276 AVLTSIQPELGHAFIDNVRVLPLELDQIIRLPFNFTSTPQETLFSIRHFDLVE 328
      DB 345 agllssfgelqyclsekkpdllelektalqnyttvtefqplvyvaesndake 397

```

OS	Homo sapiens.
XX	
PN	US6022950-A.
PN	
PD	08-FEB-2000.
XX	
PF	07-JUN-1995; 95US-0479510.
XX	
PR	07-JUN-1984; 84US-0618199.
PR	27-JUN-1991; 91US-0722484.
PR	25-APR-1985; 85US-0726808.
PR	07-JUN-1985; 85US-0742554.
PR	22-DEC-1989; 89US-0456095.
PR	14-JUN-1990; 90US-0538276.
PR	04-AUG-1993; 93US-0102387.
XX	(SERA-) SERAGEN INC.
XX	
XX	Murphy JR;
XX	
FI	WPI; 2000-160390/14.
DR	N-PSDB; AAZ90020.
DR	
XX	
PT	New two-part hybrid protein comprising a translocation domain and a
PT	cell-binding domain, for treating genetic deficiency diseases, cancer
PT	and HIV infections.
XX	
PS	Example 5; Fig 13; 32pp; English.
XX	
XX	This sequence represents the phenylalanine hydroxylase protein sequence.
CC	The encoded protein can be included in the hybrid protein of the
CC	invention and used to destroy or modify the cell that the hybrid protein
CC	is targeted to. The hybrid protein comprises a first part which is a
CC	portion of the binding domain of a cell-binding ligand, effective to
CC	cause the hybrid molecule to bind to a cell of an animal. The second
CC	part comprises a portion of a translocation domain of a naturally
CC	occurring protein (e.g. the translocation domain of diphtheria toxin) the
CC	second part translocates the third part across the cytoplasmic membrane
CC	and into the cytosol of the cell. The third part comprises a chemical
CC	entity to be introduced into the cell, where each of the first and third
CC	part is non-native with respect to naturally occurring protein, and the
CC	covalent bond attaching the second and third part is cleavable. The
CC	phenylalanine hydroxylase protein can form part of the third portion of
CC	the hybrid protein, the cell binding domain binds to a specific cell and
CC	the translocation domain transfers the hybrid molecule across the cell
CC	membrane into the cytosol. The third part of the protein, linked to the
CC	translocation domain through a cleavable bond, can then carry out its
CC	function. The hybrid molecules are useful for treating genetic deficiency
CC	diseases by delivering to affected cells an enzyme supplying the missing
CC	function, to supplement cellular levels of a particular enzyme or a
CC	scarce precursor or cofactor, to direct toxins or other poisons to
CC	destroy particular cells (such as adipocytes, cancer cells, or
CC	virus-infected cells), and to counteract viral infections such as HIV by
CC	introducing into appropriate cells antibodies to viral proteins.
XX	
SQ	Sequence 452 AA;

RESULT 9  
AAAY78593  
ID AAY78593 standard: protein: 452 AA.

XX  
AC  
AA78593:

XX  
DT 05-MAY-2000 (first entry)

XX DE Human phenylalanine hydroxylase protein sequence.

xx Phenylalanine hydroxylase; translocation domain; cell destruction;  
KW cell binding domain; genetic deficiency disease; cell targeting; cancer;  
KW adipocyte; enzyme delivery; anti-viral; HIV.

XX

```

QY 276 AVLISSPOELGHAFIDNVRVLPLELDQIIRLPENTSTPOETLFSIRHDELVE 328
  ||||| : : : ||||| : : : : : : : : : : : : : : : : : :
Db 345 agllsfgeclgycisxpkllplelektalqyvtetfoplyyvaesfndake 397

RESULT 10
AAB1327
ID AAB13327 standard; Protein: 532 AA.
AC AAB13327;
XX
XX 12-JAN-2001 (first entry)
DT
DE Caenorhabditis elegans Cod-5.
DE
DE Caenorhabditis elegans; cod-5; daf-18; insulin signalling pathway;
KW daf-2; age-1; insulin receptor; PI 3-kinase; PKB kinase;
KW PTEN lipid phosphatase; antidiabetic; anorectic; obesity; diabetes.
XX
XX Caenorhabditis elegans.
OS
OS W0200033068-A1.
XX
XX 08-JUN-2000.
PD
XX
XX 02-DEC-1999; 99WO-US28529.
PF
XX
XX 03-DEC-1998; 98US-0205658.
PR
XX
XX (GEO ) GEN HOSPITAL CORP.
PA
XX
XX Ruvkun G, Ogg S;
PI
XX
XX WPI: 2000-423022/36.
DR
XX
XX N-PSDB; AAN91624.
DR
XX
XX Diagnosing and treating obesity and impaired glucose tolerance using
PT modulators of daf-18 expression and/or activity -
PS
XX
XX Disclosure; Fig 42; 402pp; English.
XX
XX The present sequence is Cod-5 from Caenorhabditis elegans. Cod-5
XX is the aromatic amino acid hydroxylase that synthesises serotonin from
XX precursor L tryptophan. The cod-5 gene was knocked out to produce mutants
XX completely lacking in serotonin. These mutants were found to have defects
XX in metabolic control. A number of C. elegans proteins that have
XX mammalian homologues acting in the insulin signalling pathway were also
XX identified. The C. elegans age-1 gene encodes a homologue of the
XX mammalian PI 3-kinase whilst daf-2 encodes a homologue of the mammalian
XX insulin receptor. The C. elegans AKT kinase and PKB kinase act downstream
XX of daf-2 and age-1, just as their mammalian homologues act downstream of
XX insulin signalling. The C. elegans PTEN lipid phosphatase homologue,
XX DAF-18, has been found to act upstream of AKT in the pathway. This
XX discovery has enabled mammalian PTEN action to be mapped to the insulin
XX signalling pathway. Conserved DAF motifs can be used to design probes to
XX identify mammalian DAF homologues and thus to identify individuals with a
XX predisposition towards the development of glucose intolerance conditions,
XX such as obesity and diabetes.
XX
XX Sequence 532 AA:

Query Match
Best Local Similarity 10.98; Score 206.5; DB 21; Length 532;
Matches 70; Conservative 47; Mismatches 116; Indels 45; Gaps 8;

QY 104 RNW---YKLSSRFSLKWSKCPFFLDYLEAFGLLSDFLDH-----QAVIKFFELETHF 155
  ||||| : : : ||||| : : : : : : : : : : : : : : : : : :
Db 251 rktwgiyrlr---elkhkackqldnfellerhcgysennipqledickfiktgtf 307
  ||||| : : : ||||| : : : : : : : : : : : : : : : : : :
QY 156 SYYPVSGFVAPHQYLSLQDRYFIASVMKTLDKNFSLTPLDHLGLGHVPWLLHPSFS 215
  ||||| : : : ||||| : : : : : : : : : : : : : : : : : :

```

```

Db 308 rrvpvgayisardflaglayrvffctqyvrhpadpfytpepdtvhelmgjmalfadpdfa 367
QY 216 EFFINNGRLFTVKIEKVOALPSKKORIOTLQSNLIAIVRCFWFTVESGLI----- 265
  ||||| : : : ||||| : : : : : : : : : : : : : : : : : :
Db 368 qfsqeg-----laslgaseedlkkiatl-----yffisiefglssdaadspvk 411
QY 266 ---ENHEGRKAYCAVLISPOELGHAFIDNVRVLPLELDQIIRLPENTSTPOETLFSIRH 322
  ||||| : : : ||||| : : : : : : : : : : : : : : : : : :
Db 412 engsnherfkvygagllssagelqhavegsatirfdprvveqecclitfqsayfyrn 471
QY 323 FDELVELTSKLEWMLDQGLLESIPLYNOERYLSGFVFL 360
  ||||| : : : ||||| : : : : : : : : : : : : : : : : : :
Db 472 feeaqg---klrmfCnmakrpfivryn--pytesvevl 504

RESULT 11
AAW69848
ID AAW69848 standard; Protein: 924 AA.
XX
XX AAW69848;
XX
XX 07-OCT-1998 (first entry)
DT
DE Amino acid sequence of p102, a protein present in SA-17S complex.
DE
DE p102 protein; secretion associated 17S complex; SA-17S;
KW syntaxin-containing complex; SC complex; screening; modulate;
KW vesicular release; synaptic transmission; secretory process;
KW treatment; affective disorder; depression; manic-depressive disorder;
KW anxiety disorder; neurodegenerative disease; schizophrenia; anaesthesia;
KW hormonal imbalance; antigen processing; ss.
XX
XX Rattus sp.
OS
XX
XX W09828419-A2.
DR
XX
XX 02-JUL-1998.
PD
XX
XX 19-DEC-1997; 97WO-US23498.
PF
XX
XX 20-DEC-1996; 96US-0033905.
PR
XX
XX (STRD ) UNIV LELAND STANFORD JUNIOR.
PA
XX
XX Scheller RH;
PI
XX
XX WPI: 1998-377650/32.
DR
XX
XX N-PSDB; AAV42653.
DR
XX
XX New isolated vesicle secretion associated poly:peptide(s) - used to
XX develop products for treating e.g. effective disorder.
XX neurodegenerative disease, hormone imbalances, immune system
XX disorders or tumours
XX
XX Claim 1e; Pages 116-118, 133pp; English.
XX
XX The present sequence represents a p102 protein. This protein is present
XX the secretion associated 17S (SA-17S) complex. Eight proteins form the
XX SA-17S complex, which binds a syntaxin-containing (SC) complex. The
XX SA-17S polypeptides and nucleotide sequences encoding them can be used
XX for screening for compounds which modulate vesicular release involved in
XX synaptic transmission and other secretory processes. Compounds which
XX enhance binding between the SA-17S and SC complexes may be used to treat
XX an affective disorder such as depression, manic-depressive disorders and
XX anxiety disorders, or a neurodegenerative disease such as Parkinson's
XX disease or Huntington's disease. Compounds which inhibit binding between
XX the SA-17S and SC complexes may be used to treat a disorder of thought,
XX such as schizophrenia, or for anaesthesia. The compounds can also be used
XX to intervene in the endocrine system for treatment of hormonal
XX imbalances, the immune system for intervention in antigen processing,
XX secreted immunomodulators, and viral processing, as well as
XX anti-tumour applications, such as regulation of membrane trafficking
XX during rapid cell division.

```

```
XX SQ Sequence 924 AA;
Query Match 5.2%; Score 99; DB 19; Length 924;
Best Local Similarity 19.9%; Pred. No. 0.46;
Matches 90; Conservative 77; Mismatches 122; Indels 164; Gaps 27;
QY 19 KLRQLSLFFQNSQSQRAYSTPYRYRILOKNEKQALAHKHCISILEFFKNLLFVH 78
Db 233 kvegsmtqklen--vlnrasntadtlfgevlgrtkk---adstrnalnvlgrfk-flfnl 286
QY 79 LLSLKNQREGCSTDMVAVSTPFNNRLWRLSSRF-----SLWKSVCPRFFLDYLEAF 133
Db 287 plnkrniqkg---dydvind-----yekaksifgktevqvkky-----yavee 329
QY 134 GLLSDFLDHQAVIRKFELETHFSYYPVSGFV-----APHQY-LSLLQ 174
330 agiedl--reillkk1-letpstlhdqkryrlysdhagdpawqcgagqhkwtlkmq 386
QY 175 D-----RYFFIASVMRTLDKDNFSLTDLIHDILGHV----- 206
Db 387 dckeghmkslkgngpphspm-----ldldn-----darpvlgihstaslkrgssfqg 435
QY 207 ----PWLH-----PSFEFFINM--GRFTKVIK-----VOALPSK 238
Db 436 grddtwrykthprvafvekltkvlslqplnfwklwisvngslfsetaeksgqierknv 495
QY 239 KOR-----IOTLOSNIATVR--CFWFTVESGLIENHEGRKAYGAVLIS---SPOELG 286
Db 496 rdgrndfkmiqemhslvklirgallpfsalreg-----dgrq-yggvevqaelsgqqla 549
QY 287 HAFIDNVRV-----LPLELDQIIRLPENTSTPOETLFSIR-----HFDELV 327
Db 550 hv-igtirlyesitaleipndmlqii-----gdilrlrhcmvltlghtaeiek 599
QY 328 ELTSKLEWMLDQGLESTPLYNODKLSGPEVL 360
Db 600 rlaekedwlvndegltalpcqfeqsvhslqsl 632
RESULT 12
AAW19604
XX ID AAW19604 standard; Protein; 1024 AA.
XX AC AAW19604;
XX D 21-AUG-1997 (first entry)
XX DE Mycoplasma genitalium 116 kDa protein MG075 useful in vaccine.
XX KW Mycoplasma; immunogen; vaccine; diagnosis; pneumonia; inflammation.
XX OS Mycoplasma genitalium.
XX PN WO9721727-A1.
XX PD 19-JUN-1997.
XX PF 13-DEC-1996; 96WO-AU00803.
XX PR 13-DEC-1995; 95AU-0007127.
XX PA (UYME ) UNIV MELBOURNE.
XX PI Browning GF, Duffy MF, Walker ID, Whitchear KG;
XX DR WPI; 1997-332722/30.
XX PT New immunogenic polypeptide(s) from Mycoplasma species - useful in
XX vaccines and for diagnosis of Mycoplasma infection
XX PS Claim 19; Page 85-89; 110pp; English.
```

```
XX CC Isolated or recombinant immunogenic polypeptides from Mycoplasma
CC genitalium have mol.wt. of 16 kDa (AAW19603) (MG074) and 116 kDa
CC (AAW19604) (MG075). They are homologues of 16 and 116 kDa proteins
CC (see also AAW19601-02) obtd. from Mycoplasma pneumoniae. A genomic
CC DNA sequence of M. genitalium contains contiguous open reading
CC frames that code for the 2 polypeptides. Mycoplasma 16 or 116 kDa
CC proteins, or immunogenic fragments that include a T or B cell
CC epitope, can be used in vaccines for prevention and treatment of
CC Mycoplasma infections, partic. in humans. They can also be used
CC diagnostically to detect Mycoplasma, or to raise antibodies useful
CC in immunoassays.
XX SQ Sequence 1024 AA;
Query Match 5.2%; Score 98.5; DB 18; Length 1024;
Best Local Similarity 19.9%; Pred. No. 0.6;
Matches 68; Conservative 41; Mismatches 134; Indels 99; Gaps 11;
QY 29 QNSQSQRAYSTPYRYRILOKNEK---QALAHKHCISILEFFKNLLFVHLLSKN 85
Db 564 qtdsdlknlfsv-----igdlsetcnvknkltlhavknnellisvetastkikhl----- 613
QY 86 QREGCSTDMVAVSTPFNNRLWRLSSRFSLWKSVCPRFFLDYLEAFGLLSDFLDHQAV 145
Db 614 -----nvqykviavdkfelknsfik-----ellnffpdkdltpt 647
QY 146 IK--FFELETH-----FSYYPVSGFVAPHQYLSLLQDRYFFPIASVMRTLDKDNF 192
Db 648 ikkvlfesenyktrkyenegfpgyhwakfivpgtfnsaentfysal-----dkt 698
QY 193 SLTPDLIHDILGHVPMLLHPSPSEFFINMGRFLT-----KVIEKVAQALPSKKQRIQT 244
Db 699 ksirdlfadml-fgkslesvndsdfikingsftlittkndnlnllpnyslittknvgvqi 757
QY 245 LOSNLIAIVRCFWFTVESGLIENHEGRKAYGAVLISSPQELG----- 286
Db 758 vvnvfhidarilltaeqntvfsnpk-----pvikspvelskslfkwkifensvni 810
QY 287 ---HAFIDNVRVLPLELDQIIRLPENTSTPOETLFSIRHPD 324
Db 811 lkkeytkdnlkffpkadgssrleldskpdrvipfafvd 852
RESULT 13
AAW19604
XX ID AAW19604 standard; Protein; 439 AA.
XX AC AAW19604;
XX D 13-SEP-1999 (first entry)
XX DE C. pneumoniae protein involved in metabolism of nucleic acids.
XX KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
XX sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
XX vaccine; neutralising epitope.
XX OS Chlamydia pneumoniae.
XX PN WO9927105-A2.
XX PD 03-JUN-1999.
XX PF 20-NOV-1998; 98WO-IB01890.
XX PR 04-NOV-1998; 98US-0107078.
XX PD 21-NOV-1997; 97PR-0014673.
XX PA (GEST ) GENSET.
XX PI Griffiths R;
```



PR 13-JUL-1999; 99US-0143542.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 17-JUL-1999; 99US-0144086.  
 PR 18-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 20-JUL-1999; 99US-0144352.  
 PR 20-JUL-1999; 99US-0144632.  
 PR 21-JUL-1999; 99US-0144884.  
 PR 21-JUL-1999; 99US-0145218.  
 PR 21-JUL-1999; 99US-0145086.  
 PR 22-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 22-JUL-1999; 99US-0145089.  
 PR 23-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145218.  
 PR 23-JUL-1999; 99US-0145224.  
 PR 26-JUL-1999; 99US-0145276.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 28-JUL-1999; 99US-0145919.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147302.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148365.  
 PR 16-AUG-1999; 99US-0148684.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 30-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.

PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 18-OCT-1999; 99US-0159638.  
 PR 21-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 26-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match 5.0%; Score 95; DB 21; Length 920;  
 Best Local Similarity 21.7%; Pred No. 1, 2; Mismatches 174; Gaps 23;  
 Matches 97; Conservative 48;

QY 9 DPXYILKIALKRLQSLSLFQNSOSLORA-----YSTPYSYRILLOKENEKQALAR 61  
 Db 371 npaivqvvdelspnsfrifwesqkfeqtdkaepwntayslekltss----- 419  
 QY 62 HKCISILEFFKNLLFVHL-----LSLSKNQREGCSTDMVAVSTPFFNRLMYR 109  
 Db 420 ----tlqewvqsapdvhlhlpapnvpfiptdls-l-kaaddketvpllrktpf--srlwyk 472  
 QY 110 LLSRFSRLWSY-----Cp-----RFFLDYLEAF----- 133  
 Db 473 -pdtmfakpkayvkmfncplavsspdavltidfrlmdylneyayyaqvaglyygs 531  
 QY 134 -----GLLSDF---LDHOAVIKFFFELETHFSYYPVSGFA 165  
 Db 532 lsdngfeltllygnhkirillletvvgkianfevkpdrfavik----etvkeyqnykfrq 587  
 QY 166 p-HO---YLSL-LQDRYFPIASVMRTLDKDNFSLTPDLIHLLGH-----VPMLLH 211  
 Db 588 pyhaamyycslilqdtwp-----wteel--dvshleaedvakfvpmls 631  
 QY 212 PSSEPEI-----NMGRUFTKVIKV-----QALPSKQRIOTLOSMLIAIVRC 255  
 Db 632 rtflecyiagnvenneasmvkhiedvifndpkpicrplfps-----ghlnrvvki--- 683  
 QY 256 FWTVESGLIENHEGRKAYCAVLISSPQELGHAFIDNVRLPDLPLEDOIIRLP-FWTSTPQ 314  
 Db 684 -----gegmkyfyhqdgdnpsdensalvhyiqvhrddfsmniklqifglvakq 731  
 QY 315 ETFSIRHFDELVELTSKLEWMLDOGL 341  
 Db 732 atfhqrlrtveqlgyita-laqrndagi 757

RESULT 15



AAG31678  
ID AAG31678 standard; Protein; 944 AA.  
XX AAG31678;  
XX  
XX  
DT 17-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 380B3.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
XX Arabidopsis thaliana.  
XX  
XX EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130449.  
XX 23-APR-1999; 99US-0130510.  
XX 28-APR-1999; 99US-0130891.  
XX 30-APR-1999; 99US-0131148.  
XX 30-APR-1999; 99US-0132048.  
XX 04-MAY-1999; 99US-0132407.  
XX 05-MAY-1999; 99US-0132484.  
XX 06-MAY-1999; 99US-0132485.  
XX 06-MAY-1999; 99US-0132486.  
XX 07-MAY-1999; 99US-0132487.  
XX 11-MAY-1999; 99US-0132863.  
XX 14-MAY-1999; 99US-0134256.  
XX 14-MAY-1999; 99US-0134218.  
XX 14-MAY-1999; 99US-0134219.  
XX 14-MAY-1999; 99US-0134321.  
XX 14-MAY-1999; 99US-0134370.  
XX 18-MAY-1999; 99US-0134768.  
XX 19-MAY-1999; 99US-0134941.  
XX 20-MAY-1999; 99US-0135124.  
XX 21-MAY-1999; 99US-0135353.  
XX 24-MAY-1999; 99US-0135629.  
XX 25-MAY-1999; 99US-0136021.  
XX 27-MAY-1999; 99US-0136392.  
XX 28-MAY-1999; 99US-0136782.  
XX 01-JUN-1999; 99US-0137222.  
XX 03-JUN-1999; 99US-0137328.  
XX 04-JUN-1999; 99US-0137502.  
XX 07-JUN-1999; 99US-0137724.  
XX 08-JUN-1999; 99US-0138094.  
XX 10-JUN-1999; 99US-0138540.  
XX 10-JUN-1999; 99US-0138847.  
XX 14-JUN-1999; 99US-0139119.  
XX 16-JUN-1999; 99US-0139452.  
XX 16-JUN-1999; 99US-0139453.  
XX 17-JUN-1999; 99US-0139492.  
XX 18-JUN-1999; 99US-0139454.  
XX 18-JUN-1999; 99US-0139455.  
XX 18-JUN-1999; 99US-0139456.  
XX 18-JUN-1999; 99US-0139457.  
XX 18-JUN-1999; 99US-0139458.  
XX 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 28-JUN-1999; 99US-0140961.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142134.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.

```

PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151203.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154779.
PR 20-SEP-1999; 99US-0155139.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 21-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161820.
PR 28-OCT-1999; 99US-0161932.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match
Best Local Similarity 5.0%; Score 95; DB 21; Length 944;
Matches 97; Conservative 48; Mismatches 128; Indels 174; Gaps 23;

QY 9 DPYKILKLRQSLRSLFQNSQLQRA-----YSTPYSYVYRIILQKENKEKQALAR 61
DB 395 npaivqkvvdelspsnfrifwesqkfgtdkaepwytatslekitts----- 443
QY 62 HKCISILEFFKNLLFVHL-----LSLSKNREGCSTDMVYSTPPFNRLWYR 109
DB 444 ----tiqewqsapdvhlhlpapnfvipdtsl-kdaddketvpllrktpf--srlwyk 496
QY 110 LLSRSFSLWKSX-----CP-----RFELDYLEAF----- 133
DB 497 -pdtmfskpavkmdfncplavssdaavltidfrllmdylneyayyaqvaqlyygs 555
QY 134 -----GLLSDF-----LDHQAVIKFEELETHFSYYPVSGEVA 165

```

```

DB 556 lsdngfeltllgynhklrillevtvvgkianfevkpdrfavik-----etvtkeygnykfrq 611
QY 166 P-HQ---YLSL-LQDRYFPIASVMRTLDKDNFSLTPDLIHLLGH-----VPWLLH 211
DB 612 pyhqamyycslilqdtwp-----wteel--dvshleaedvakfvpmlis 655
QY 212 PSTSEFFI-----NMGRLETKVIEKV-----QALPSKKQRIOTLOSNIIVRC 255
DB 656 rtflecyiagnvenneasmvkhiedvifndpkpicrplfps-----qhltnrvvki--- 707
QY 256 FWFTVESGLIENHEGRKAYCAVLISSPQELGHAFIDNVRYLPLELDQIIRLP-FNTSTPQ 314
DB 708 -----gegmkyfyhqdgnspsdensalvhyiqvhrddfsmnkiklqifglvakq 755
QY 315 ETLSIRHFEDELVELTSKLEWMLDOGL 341
DB 756 atfhqlrtveqlgyita-laqrndsgi 781

```

Search completed: June 29, 2001, 13:52:48  
 Job time: 7960 sec



```
34 euGlnArgAlaTyrSerThrProTyrSerTyrTyrArgIleIleLeuGln 50
1210726TCCACAGTGCATACGACCCCATATCTCTACTACCGAATCATCTACAA 1210677
51 LysGluAsnLysClnLysClnAlaLeuAlaArgHisLysCysIleSerIle 67
1210677GAGGAAATTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1210627
67 eLeuGluPhePheLysAsnLeuLeuPheValHisLeuLeuSerLeuSerL 84
1210626TTTAGAATTTTCAAAAGCTTACTCTTTGTTTCATCTCTCTGTCATTACAA 1210577
84 ysAsnGlnArgGluGlyCysSerThrAspMetAlaValValSerThrPro 100
1210577GAGAAATCAAAAGGGAAGGTTGCTCCACATGATATGGCTGTTGTAAGCACTCC 1210527
101 PheAsnArgAsnLeuTyrPyrArgLeuLeuSerSerArgPheSerLe 117
120526TTTTTAATCGGAATTTATGTTATGCTACTCTCTTTCCTCAGGGTTTCTCT 1210477
117 uTrpLysSerTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheG 134
1210476ATGGAAGAGCTATTGTCAGAGATTTTCTT. GATTACTTAGAAGCTTTCG 1210428
134 LysLeuLeuSerAspPheLeuAspHisGlnAlaValIleLysPheGlu 150
1210427GTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1210378
151 LeuGluThrHisPheSerTyrTyrProValSerGlyPheValAlaProH 167
1210377TTAGAAACACATTTTTCCTATTTCCGCTTTCAGAGATTTGAGTCTCCCA 1210328
167 sGlnTyrLeuSerLeuLeuGlnAspArgTyrPheProIleAlaSerValM 184
1210327TCAATACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1210278
184 etArgThrLeuAspLysAspAsnPheSerLeuThrProAspLeuIleHis 200
1210277TGGGAATCTCTGATTAAGATATTTCTCTTAACTCTGATCTCATCCAT 1210228
201 AspLeuLeuGlyHisValProTrpLeuLeuHisProSerPheSerGluPh 217
1210227GACCTTTTAGGGCAGCTGCTTGGCTTCTACATCCCTCATTTCTGAAT 1210178
217 ePheIleAsnMetGlyArgLeuPheThrLysValIleGluLysValGln 234
1210177TTTCATAAACATGGGAAGACTCTTCACTAAAGTCATAGAAAAAGTACAAG 1210128
234 laLeuProSerLysLysGlnArgIleGlnThrLeuGlnSerAsnLeuIle 250
1210127CTCTCTAGTAAACAAACAGCATACAAACCTTACAAAGCATCTGATC 1210078
251 AlaIleValArgCysPheThrProPheThrValGluSerGlyLeuIleGlu 267
1210077GCTATTTGACGCTGCTTTGGTTTACTCTTGAAGACGAGCATTTATGAAA 1210028
267 nHisGluGlyArgLysAlaTyrGlyAlaValLeuIleSerSerProGlnG 284
1210027CCATGAAGGAAGAAAGCATATGAGCCGCTTCTATCAGTCTCCCTCAGG 1209978
284 luLeuGlyHisAlaPheLeuAspAsnValArgValLeuProLeuGluLeu 300
1209977AACTTGACACGCTTCTATTTGATACAGTACGCTGCTCTCCCTTTAGAATTG 1209928
301 AspGlnIleIleArgLeuProPheAsnThrSerThrProGlnGluThrLe 317
1209927GATCAGATTTATGCTCTTCCCTTCAATACATCACTCCCAAGAGACTTT 1209878
317 uPheSerIleArgHisPheAspGluLeuValGluLeuThrSerLysLeuG 334
1209877ATTTTCAATAAGACATTTTGTAGAACTGGTGAACCTCACTTCAAAATTAG 1209828
```

```
334 luTrpMetLeuAspGlnGlyLeuLeuGluSerIleProLeuTyrAsnGln 350
1209827AATGGATGCTCGACCAAGTCTGTAGAACATCAATCCCTTTACAAATCAA 1209778
351 GluLysTyrLeuSerGlyPheGluValLeuCysGln 362
1209777GAGAAATATCTTCTGTTTGGAGTACTTTGGCCAA 1209742
seq_name: /SIDS1/gcdata/geneseq/geneseq/NA1997.DAT:AAAT62529
seq_documentation_block:
ID AAAT62529 standard; DNA; 1030 BP.
XX
AC AAAT62529;
XX
DT 06-JUN-1997 (first entry)
XX
DE Tyrosine hydroxylase truncated DNA construct rTHdel.
XX
KW Analgesic; pain; bioartificial organ; tyrosine hydroxylase;
KW norepinephrine; rTHdel; catecholamine; ss.
XX
OS Rattus sp.
XX
Key Location/Qualifiers
FH 1..6
FT 5'UTR
FT /*tag= a
FT 1..1030
FT exon
FT /*tag= b
FT /codon_start= 7
FT 3'UTR
FT 1018..1030
FT /*Lag= c
XX
PN WO9640959-A1.
XX
PD 19-DEC-1996.
XX
PF 07-JUN-1996; 96MO-US09629.
XX
PR 07-JUN-1995; 95US-0481917.
XX
PA (CYTO-) CYTOTHERAPEUTICS INC.
XX
PI Saydoff J, Wong S;
XX
WPI; 1997-087062/08.
XX
Stably transformed cells expressing endorphin, enkephalin and
catecholamine - and artificial organs contg. them, useful for
control of pain, esp. implanted in the CNS
XX
Example; Page 66; 114pp; English.
XX
A DNA sequence (AAAT62529), designated rTHdel, codes for truncated rat
tyrosine hydroxylase. It was generated by PCR amplification (see
also AAAT62526-27) of TH cDNA and insertion of the PCR product into
vector pCDNA3. rTHdel can be utilised in the development of
truncated TH in e.g. rat pancreatic endocrine RIN host cells.
XX
are capable of norepinephrine synthesis. Sequential transfection
of host cells with diff. vectors, or with a polyclonal vector,
allows produ. of cell lines that produce more than one analgesic
cpd. Such cell lines can be encapsulated to form bioartificial
organs that can be implanted e.g. in the CNS for the control of
pain.
XX
Sequence 1030 BP; 223 A; 301 C; 274 G; 232 T; 0 other;
XX
alignment_scores:
Quality: 223.00 Length: 243
Ratio: 1.640 Gaps: 5
```

Percent Similarity: 55.967 Percent Identity: 26.749

alignment block:

US-09-438-185-1047 x AAT62529

Align seg 1/1 to: AAT62529 from: 1 to: 1030

[illegible]

seq\_name: /SIDS1/gcgdata/geneseq/geneseq/NA1997.DAT:AAT62530

seq\_documentation\_block:

ID AAT62530 standard; DNA; 1037 BP.

XX

AC AAT62530;

XX	06-JUN-1997	(first entry)
DT		
XX	Tyrosine hydroxylase truncated DNA construct rTHdELKS.	
DE		
XX	Analgescic; pain; bioartificial organ; tyrosine hydroxylase;	
KW	norepinephrine; catecholamine; rTHdELKS; ss.	
KW		
XX		
OS	Rattus sp.	
OS		
Key	Location/Qualifiers	
FF	1..13	
FT	/*tag= a	
FT	1..1037	
FT	/*tag= b	
FT	/codon_start= 14	
FT	1025..1037	
FT	/*tag= c	
FT		
XX		
XX	WO9640959-A1.	
PN		
XX		
FD	19-DEC-1996.	
XX		
XX	07-JUN-1996; 96WO-US09629.	
PF		
XX		
XX	07-JUN-1995; 95US-0481917.	
PR		
XX		
XX	(CYTO-) CYTOTHERAPEUTICS INC.	
PA		
XX		
XX	Saydoff J, Wong S;	
PI		
XX	WPI; 1997-087062/08.	
DR		
XX		
XX	Stably transformed cells expressing endorphin, enkephalin and	
PT	catecholamine - and artificial organs contg. them, useful for	
PT	control of pain, esp. implanted in the CNS	
PT		
XX		
XX	Example; Page 67-68; 114pp; English.	
PS		
XX		
XX	A DNA sequence (AAT62530), designated rTHdELKS, codes for truncated	
CC	rat tyrosine hydroxylase and is designed to optimise translation	
CC	efficiency by placing the consensus Kozak sequence immediately	
CC	upstream of the start codon. It was generated from vector	
CC	pCDNA3-rTHdEL-45 by PCR amplification (see also AAT62527-28).	
CC	rTHdELKS can be utilised in the development of vectors (see also	
CC	AAT62536) for expression of truncated rTH in e.g. rat pancreatic	
CC	endorine RIN host cells. Such cells, pref. also contg. dopamine	
CC	beta-hydroxylase sequences, are capable of norepinephrine	
CC	synthesis. Sequential transformation of host cells with diff.	
CC	vectors, or with a polycistronic vector, allows prodn. of cell	
CC	lines that produce more than one analgesic cpd. Such cell lines	
CC	can be encapsulated to form bioartificial organs that can be	
CC	implanted e.g. in the CNS for the control of pain.	
CC		
XX		
XX	Sequence 1037 BP; 224 A; 306 C; 275 G; 232 T; 0 other	
SO		

alignment scores:

Cores:  
Quality: 223.00  
Length: 243

Quality:	2.5:00
Ratio:	1.640
Gaps:	5

Percent Similarity:	55.967	Percent Identity:	26.749
---------------------	--------	-------------------	--------

alignment block:

alignment\_block;  
US-09-438-185-1047 x AAT62530

Align seg 1/1 to: AAT62530 from: 1 to: 1037

112 serSerArgPheSerLeuTrpLysSer..... 120

```
serseAr gneserLeurIprysseI..... |||:::
      ::::::::::::::::::::
```

206 GCGGAAGAGATTGCTACCTGGAAGGAGGTATATGTACGCTGAAGGGCCT 255

[illegible]

121 .TyrCysProArgpheLeuAspTyrLeuGluAlapheGlyLeuLeuS 137

```

::||| | ||:|||||::: :: :||
::||| | ||:|||||::: :: :||
- - - - - Fourscore sixtyfour'

```

— — — — —



```
1273 CGGAGACAGATTGCTACCTACCTGGAGGAGGTATATGTACGCTGAAGGGCCT 1322
121  TyrCysProArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeu 137
1323 CTATGCTACCATGTCGCTGCGGGAGACCTGGAGGTTCCACGCTTCTGG 1372
137  erAspPheLeuAspHis.....GlnAlaVal 145
1373 AACGGTACTGTGCTACCGAGAGGAGACACATCCACAGCTGGAGGACGTG 1422
146  IleLysPhePheGluLeuGluThrHisPheSerTyrTrpProValSerG 162
1423 TCCGGCTCTCTGAAGAGCGGACTGGCTTCCAGCTGGAGCCGCTGGCGG 1472
162  yPheValAlaProHisGlnTyrLeuSerLeuLeuGlnAspArgTyrPheP 179
1473 TCTACTGTCCGCCCGGATTTCTGCCAGCTGGCTTCCGGCTGCTTTC 1522
179  rolleAlaSerValMetArgThrLeuAspLysAspAsnPheSerLeuThr 195
1523 AATGCACCCAGTATATCCGCCATGCTCTCCATCATGCTACCTGAG 1572
196  ProAspLeuIleHisAspLeuLeuGlyHisValProTrpLeuLeuHisPr 212
1573 CCGGACGTGCTGATGAGCTGTGGAGCATGTACCCATGTGGCTGACCG 1622
212  oSerPheSerGluPhePheIleAsnMetGlyArgLeuPheThrLysValI 229
1623 CACATTTGGCCAGTTCTCCAGGACATTTGGA..... 1653
229  leGluLysValGlnAlaLeuProSerLysLysGlnArgIleGlnThrLeu 245
1654 .....CTTGCACTCTCTGGGGGCTCAGATGAAGAAATTTGAAACATC 1695
246  GlnSerAsnLeuIleAlaIleValArgCysPheTrpPheThrValGluSe 262
1696 TCCACG.....GTGTACTGTTCACCTGTGCAATT 1724
262  rGlyLeuIleGluAsnHisGluGlyArgLysAlaTyrGlyAlaValLeuI 279
1725 CGGCTATGTAACAGAAATGGGAGCTGAAGCTTATGGTGCAGGCTCC 1774
279  leSerSerProGlnGluLeuGlyHisAlaPheIleAspAsnValArgVal 295
1775 TGTTCTCTACGAGAGCTCTGCACTCCCTGTCAGAGGAGCTGAGCTC 1824
296  LeuProLeuLeuAspGlnIleIleArgLeuProPheAsnThrSerTh 312
1825 CGAGCTTTGACCCAGACACAGCAGCTGTGAGCCCTACCAAGATCAAA 1874
312  rProGlnGluThrLeuPheSerIleArgHisPhe.....AspG 325
1875 CTACAGCCTGTGTACTTGTGTCGAGAGCTTCAATGACGCCAAGGACA 1924
325  luLeuValGluLeuThrSerLysLeuGlu 334
1925 AGCTCAGGAAGTATGCTCTGCTATCCAG 1953
seq_name: /SIDS1/gcgdata/geneseq/geneseqn/NA1997.DAT:AAAT62548
seq_documentation_block:
ID AAAT62548 standard; DNA; 5540 BP.
XX
AC AAAT62548;
XX
DT 07-JUN-1997 (first entry)
XX
DE IgSP-hPOMCdelACTH-IREs-rfidel-IREs-bDBH-IREs-Zeocin-073 DNA.
XX
KW Analgesic; pain; bioartificial organ; pro-opiomelanocortin; POMC;
KW beta-endorphin; tyrosine hydroxylase; dopamine beta-hydroxylase;
KW IgSP-hPOMC-delACTH-IREs-rfidel-IREs-bDBH-IREs-Zeocin-073;
KW internal ribosome entry site; ss.
```

```
XX OS Chimeric Homo sapiens;
OS Chimeric picornavirus;
OS Chimeric Rattus sp.;
OS Chimeric Bos taurus.
XX
FH Key Location/Qualifiers
FT 5'UTR 1..118 /*tag= a
FT exon 1..164 /*tag= b
FT FT /*codon_start= 119
FT intron 165..243 /*tag= c
FT exon 244..557 /*tag= d
FT FT /*tag= 1155
FT intron 1156..2166 /*tag= e
FT exon 2167..2766 /*tag= f
FT FT /*tag= 2767..4560
FT exon 4561..5159 /*tag= g
FT intron 5160..5540 /*tag= h
FT FT /*tag= 5535..5540
FT 3'UTR /*tag= k
XX
XX WO9640959-A1.
XX
XX 19-DEC-1996.
XX
XX 07-JUN-1996; 96WO-US09629.
XX
XX 07-JUN-1995; 95US-0481917.
XX (CYTO-) CYTOTHERAPEUTICS INC.
XX Saydoff J, Wong S;
XX WPI; 1997-087062/08.
XX
XX Stably transformed cells expressing endorphin, enkephalin and
XX catecholamine - and artificial organs contg. them, useful for
XX control of pain, esp. implanted in the CNS
XX
XX Example; Page 84-88; 114pp; English.
XX
XX A DNA sequence (AAAT62548) comprises a fusion between IgSP, a human
XX pro-opiomelanocortin (POMC) gene in which the ACTH portion has
XX been deleted (see also AAAT62524), a viral IRES, a truncated rat
XX tyrosine hydroxylase gene (see also AAAT62529), another IRES, a bovine
XX dopamine beta-hydroxylase (bDBH) (see also AAAT62535) gene and a
XX zeocin gene. Host cell lines (e.g. RIN rat pancreatic endocrine)
XX transformed with a vector carrying the construct can convert the
XX POMC-delACTH to beta-endorphin, and utilize endogenous catecholamine
XX synthetic enzymes with the recombinant TH and DBH sequences to
XX produce norepinephrine. Sequential transfection of such cells
XX with diff. vectors, or with a polyclonistic vector allows prodn. of
XX cell lines that produce more than one analgesic cpd. Such cell
XX lines can be encapsulated to form bioartificial organs that can be
XX implanted e.g. in the CNS for the control of pain.
XX
XX Sequence 5540 BP; 1132 A; 1697 C; 1587 G; 1124 T; 0 other;
```

```
alignment_scores:
Quality: 223.00 Length: 243
Ratio: 1.640 Gaps: 5
Percent Similarity: 55.967 Percent Identity: 26.749
```

## alignment\_block:

```
US-09-438-185-1047 x AAT62548
Align seg 1/1 to: AAT62548 from: 1 to: 5540
112 SerSerArgPheSerLeuTrpLysSer..... 120
1348 GCGGAGAGATTGCTACCTGGNAGGAGGTATATGTCACGCTGAAGGCGCT 1397
121 TyTCysProArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeuS 137
1398 CTATGCTACCAATGCTGCCGGGACCACTGGAGGGTTCCAGCTTCTGG 1447
137 erAspPheLeuAspHis.....GlnAlaVal 145
1448 AACGGTACTGTGGCTACCGAGAGAGACATCCACAGCTGGAGGACGTG 1497
146 IleLysPhePheGluLeuGluThrHisPheSerTyrTyrProValSerG1 162
1498 TCCGCTCTTGAAGGAGCGGACTGGCTTCCAGCTGGACCGCTGGCGG 1547
162 yPheValAlaProHisGlnTyrLeuSerLeuLeuGlnAspArgTyrPheP 179
1548 TCCTACTGTCCGCGCTGATTTCTGGCCAGCTGCTCCGCTCCGCTGTTTC 1597
179 roileAlaSerValMetArgThrLeuAspLysAspAsnPheSerLeuThr 195
1598 ATGCACCCAGTATATCCGCCATCTCTCCATCATTCACCTGAG 1647
196 ProAspLeuIleHisAspLeuGlyHisValProTrpLeuLeuHisPr 212
1648 CCGGACTGTCCATGAGCTGTGGACATGTACCATGTTGGCTGACCG 1697
212 oSerPheSerGluPhePheIleAsnMetGlyArgLeuPheThrLysValI 229
1698 CACATTGCCAGTCTCCAGGACATGGA..... 1728
229 leGluLysValGlnAlaLeuProSerLysLysGlnArgIleGlnThrLeu 245
1729 .....CTTGCATCTCTGGGGCCCTCAGATGAAGAAATTGAAAACTC 1770
246 GlnSerAsnLeuIleAlaIleValArgCysPheThrPheThrValGluSe 262
1771 TCCACG.....GTGTACTGTGTCACTGTGGAATT 1799
262 rGlyLeuIleGluAsnHisGluIleArgLysAlaTyrGlyAlaValLeuI 279
1800 CGGGCTATGTAACAGAAATGGGAGCTGAAGGCTTATGTGCAGGCTGC 1849
279 leSerSerProGlnGluLeuGlyHisAlaPheIleAspAsnValArgVal 295
1850 TGCTCTCTACGAGAGCTCTCCACTCCCTGTGCAGAGAGCCTGAGGTC 1899
296 LeuProLeuGluLeuAspGlnIleIleArgLeuProPheAsnThrSerth 312
1900 CGACCTTTGACCCAGACACAGCAGCTGTGCAGCCCTTCCAGATCAAC 1949
312 rProGlnGluThrLeuPheSerIleArgHisPhe.....AspG 325
1950 CTACACGCTGTGTACTTGTGTCCGAGAGCTTCAATGACGCCCAAGGACA 1999
325 luLeuValGluLeuThrSerLysLeuGlu 334
2000 AGCTCAGGAACCTATGCTCTCTGATCCAG 2028
seq_name: /SIDS1/gcgdata/geneseq/NA1997.DAT:AAT62535
seq_documentation_block:
ID AAT62535 standard; DNA; 3425 BP.
XX
AC AAT62535;
XX
```

```
06-JUN-1997 (first entry)
XX rTHdel-IRES-bDBH DNA sequence.
XX
XX Analgesic; pain; bioartificial organ; tyrosine hydroxylase;
KW dopamine beta-hydroxylase; internal ribosome entry site; IRES;
KW norepinephrine; catecholamine; rTHdel-IRES-bDBH; ss.
XX
XX Chimeric Rattus sp.;
OS Chimeric picornavirus;
OS Chimeric Bos taurus.
FH
FH Key Location/Qualifiers
FT 5'UTR 1..5
FT exon /*tag= a
FT 1..1017
FT intron /*tag= b
FT 1018..1617
FT exon /*tag= c
FT 1618..3425
FT /*tag= d
FT 3412..3425
FT /*tag= e
FT misc_feature 1025..1617
FT /*tag= f
FT /*product= IRES
XX
XX WO9640959-A1.
XX
XX 19-DEC-1996.
XX
XX 07-JUN-1996; 96WO-US09629.
XX
XX 07-JUN-1995; 95US-0481917.
XX (CYTO-) CYTOTHERAPEUTICS INC.
XX
XX Saydoff J, Wong S;
XX
XX WPI; 1997-087062/08.
XX
XX Stably transformed cells expressing endorphin, enkephalin and
XX catecholamine - and artificial organs contg. them, useful for
XX control of pain, esp. implanted in the CNS
XX
XX Example; Page 69-71; 114pp; English.
XX
XX 2 DNA constructs (AAT62535 and AAT62536) respectively comprise a
XX truncated rat tyrosine hydroxylase sequence, rTHdel (see also
XX AAT62529) or rTHdelKS (see also AAT62530), joined via an IRES sequence
XX to the bovine dopamine beta-hydroxylase gene. Expression of the
XX constructs in transfected RIN or AtT-20 cells complements the
XX host cells' catecholamine synthesizing enzymes, allowing prodn.
XX of norepinephrine. Sequential transfection of host cells with
XX diff. vectors or with a polycistronic vector (see also AAT62543,
XX AAT62548), allows prodn. of cell lines that produce more than one
XX analgesic cpd. Such cell lines can be encapsulated to form
XX bioartificial organs that can be implanted e.g. in the CNS for the
XX control of pain.
XX
XX Sequence 3425 BP; 687 A; 1093 C; 962 G; 683 T; 0 other;
XX
XX alignment_scores:
XX Quality: 221.00 Length: 230
XX Ratio: 1.713 Gaps: 4
XX Percent Similarity: 56.087 Percent Identity: 26.957
XX
XX alignment_block:
XX US-09-438-185-1047 x AAT62535
XX
XX Align seg 1/1 to: AAT62535 from: 1 to: 3425
```





306 AACGGTACTGTGCTACCGAGGAGACAGCATCCACAGCTGGAGACGCTG 355  
146 ILySaphePheLeuGluthrHisPheSerTyrrProValSerGI 162  
356 TCCCGGTCTTTGAAGAGCGCTGCCTTCAGCTGCCAGCACCCTGGCCGG 405  
162 yPheValAlaProHisGlnTyrrLeuSerLeuGlnAsparGlyrPheP 179  
406 TCtACTGTCCGCCGTGATTTTTCCTGGCCAGCTCGCCTCCGCGTGTTTC 455  
179 rotlealaSerValMetArgThrLeuAspLysAspAsnPheSerLeuThr 195  
456 AATGCACCCAGTATATCCGCCATCTCCTCACCTATGCTACCTGTAG 505  
196 ProAspLeuIleHisAspLeuGluGlyHisValProTrpLeuLeuHisPr 212  
506 CCGGACTGCTGCATGAGCTGTGGGACATGTACCCATGTGGCTGACCG 555  
212 oSerPheSerGluPhePheIleAsnMetGlyArgLeuPheThrLysValI 229  
556 CACATTGGCCAGTCTCCAGGACATTTGA..... 586  
229 leGluLySValGlnAlaLeuProSerLysLysGlnArlleGlnThrLeu 245  
587 .....CTTGCACTCTCTGGGGGCTCAGATGAAGAATAATGAAAACTC 628  
246 GlnSerAsnLeuIleAlaIleValArgCysPheTrpPheThrValGluSe 262  
629 TCCAGC.....GTGTACTGGTCTACTGTGGAAT 657  
262 rGLyLeuIleGluAsnHisGluGlyArgLysAlaTyrrGlyAlaValLeuI 279  
658 CGGGCTATCTAACACAGATGGGGAGCTGAAGGCTTATGGTGCAGGGCTG 707  
279 leSerSerProGlnGluLeuGlyHisAlaPheIleAspAsnValArgVal 295  
708 TGtCTTCCTACGAGAGCTCCCGCACTCCCTGTCAGAGGAGCCCTGAGGTC 757  
296 LeuProLeuGluLeuAspGlnIleIleArgLeuProPheAsnThrSerTh 312  
758 CGAGCCTTTTGACCCAGACACAGCAGCTGTGCAGGCCCTACCAAGATCAAAC 807  
312 rProGlnGluThrLeuPheSerIleArgHisPheAspGlu 325  
808 CTACCAAGCCTGTGTACTTTGTGTCGAGAGCTTCAATGAC 847  
ame: /SIDS1/gcgdata/geneseq/geneseqn/NA1994.DAT.AAQ44699

seq documentation block:  
ID AAQ44699 standard; cDNA; 1830 BP.  
XX AAQ44699;  
AC  
DT Human tyrosine hydroxylase cDNA.  
DE  
XX  
KW Human; rat; tyrosine hydroxylase; TH; substitution; N-terminal;  
KW enhanced enzymatic activity; wild-type; transformation; dopamine;  
KW Parkinson's disease; Alzheimer's disease; brain; encapsulation;  
KW selectively permeable; polymer capsule; antibody; ss.  
XX Homo sapiens.  
FH Key Location/Qualifiers  
FF 20..1513  
CDS /\*tag= a  
FT /product= Human TH  
polyA\_site 1795..1801  
FT /\*tag= b  
XX US5300436-A.

945 AGTGACCCAGATATATCGCCACCGCTCTCGCCCATGCACCTCCCTGAG 994  
196 ProAspLeuIleHisAspLeuGlyHisValProTrpLeuHisPr 212  
995 CCGGACTGTGCGACGAGCTCTGGGACAGTCCCATCTGCGCCGACG 1044  
212 oSerPheSerGluPhePheIleAsnMetGlyArgLeuPheThrLysVal 229  
1045 CACTTCGGCGAGTCTCGCAGGACATGGC.....C 1076  
229 leGluLysValGlnAlaLeuProSerLysLysGlnArgIleGlnThrLeu 245  
1077 TGGCGTCTGTAAGCAAGACGGGAGGTGAAGGCTATGTGCGCGGCTG 1126  
246 GlnSerAsnLeuIleAlaValArgCysPheTrpPheThrValGlu 262  
1127 TCA.....TGGTTCACGGTGAGATT 1146  
262 rGlyLeuIleGluAsnHisGluGlyArgLysAlaTyrGlyAlaValLeu 279  
1147 CGGCGTGTGAAGCAAGACGGGAGGTGAAGGCTATGTGCGCGGCTG 1196  
279 leSerSerProGlnGluLeuGlyHisAlaPheIleAspAsnValArgVal 295  
1197 TGTCTCTCTACGGGAGCTCTCGCAGCTGCTGTGAGGAGCTGAGATT 1246  
296 LeuProLeuGluLeuAspGlnIleIleArgLeuProPheAsnThrSerTh 312  
1247 CGGCGCTTCGACCTGAGGCTCGCGCGCTGCAGCCCTACCAAGACGAG 1296  
312 rProGlnThrLeuPheSerIleArgHisPhe.....Asp 325  
1297 GTACCACTAGTCTACTCTGCTGCTGAGAGTCTAGTGACGCCAGGAG 1346  
325 leuValGluThrSerLysLeuGlu 334  
1347 ACCTCAGGAGCTATGCTCAGCAGATCCAG 1375

seq\_name: /SIDS1/gcgdata/geneseq/geneseq/NA2000.DAT:AAZ90020

seq\_documentation\_block:

ID AAZ90020 standard; cDNA; 1650 BP.

XX AC AAZ90020;

XX 05-MAY-2000 (first entry)

XX Human phenylalanine hydroxylase cDNA sequence.

XX Phenylalanine hydroxylase; translocation domain; cell destruction;  
XX cell binding domain; genetic deficiency disease; cell targeting; cancer;  
XX adipocyte; enzyme delivery; anti-viral; HIV; ss.

XX OS Homo sapiens.

XX PN US6022950-A.

XX PD 08-FEB-2000.

XX PF 07-JUN-1995; 95US-0479510.

XX PR 07-JUN-1984; 84US-0618199.

XX PR 27-JUN-1991; 91US-0722484.

XX PR 25-APR-1985; 85US-0726808.

XX PR 07-JUN-1985; 85US-0742554.

XX PR 22-DEC-1989; 89US-0456095.

XX PR 14-JUN-1990; 90US-0538276.

XX PR 04-AUG-1993; 93US-0102387.

XX (SERA-) SERAGEN INC.

XX PA Murphy JR;

XX PI

DR WPI: 2000-160390/14.  
XX P-PSDB; AAY78593.

XX New two-part hybrid protein comprising a translocation domain and a  
PT cell-binding domain, for treating genetic deficiency diseases, cancer  
PT and HIV infections.

XX Example 5; Fig 13; 32pp; English.

XX This sequence represents the phenylalanine hydroxylase coding sequence.  
XX The encoded protein can be included in the hybrid protein of the  
XX invention and used to destroy or modify the cell that the hybrid protein  
XX is targeted to. The hybrid protein comprises a first part which is a  
XX portion of the binding domain of a cell-binding ligand, effective to  
XX cause the hybrid molecule to bind to a cell of an animal. The second  
XX part comprises a portion of a translocation domain of a naturally  
XX occurring protein (e.g. the translocation domain of diphtheria toxin) the  
XX second part translocates the third part across the cytoplasmic membrane  
XX and into the cytosol of the cell. The third part comprises a chemical  
XX entity to be introduced into the cell, where each of the first and third  
XX part is non-native with respect to naturally occurring protein, and the  
XX covalent bond attaching the second and third part is cleavable. The  
XX phenylalanine hydroxylase protein can form part of the third portion of  
XX the hybrid protein. The cell binding domain binds to a specific cell and  
XX the translocation domain transfers the hybrid molecule across the cell  
XX membrane into the cytosol. The third part of the protein, linked to the  
XX translocation domain through a cleavable bond, can then carry out its  
XX function. The hybrid molecules are useful for treating genetic deficiency  
XX diseases by delivering to affected cells an enzyme supplying the missing  
XX function, to supplement cellular levels of a particular enzyme or a  
XX scarce precursor or cofactor, to direct toxins or other poisons to  
XX destroy particular cells (such as adipocytes, cancer cells, or  
XX virus-infected cells), and to counteract viral infections such as HIV by  
XX introducing into appropriate cells antibodies to viral proteins.

XX Sequence 1650 BP; 433 A; 405 C; 386 G; 426 T; 0 other;

alignment\_scores:

Quality: 213.50 Length: 233

Ratio: 1.642 Gaps: 4

Percent Similarity: 55.794 Percent Identity: 26.609

alignment\_block:

US-09-438-185-1047 x AAZ90020 ..

Align seg 1/1 to: AAZ90020 from: 1 to: 1650

104 ArgAsnLeuTrpTyrArgLeuLeuSerSerArgPheSerLeuTrpLys 120

772 AAGAAACATGGGACAGCTGTCAAGACTCTGAAGCTCTGTATATAAC 821

120 rTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeu 137

822 CCAATGCT.....TGCTATGAGTACAATCAATTTTCCACTTCTTG 862

137 erAspPheLeuAspHis.....GlnAlaVal 145

863 AAAAGTACTGTGCTTCCATGAGATACATTTCCCGACCTGGAAGACGTT 912

146 IleLysPhePheGluLeuGluThrHisPheSerTyrTyrProValSerG 162

913 TCTCAATTCCTCGACACTTGCACCTGTGTTCGCGCTCCGACCTGTGGCTG 962

162 yPheValAlaProHisGlnTyrLeuSerLeuGlnAspArgTyrPheP 179

963 CCGCTTCTCTCTCGGAGTTCTTGGTGGCTGCGCTTCCGAGCTCTTC 1012

179 rolLeuAlaSerValMetArgThrLeuAspLysAspAsnPheSerLeuThr 195

1013 ACTGCACACAGTACATCAGACATGGATCCCAAGCCCATGTATACCCCGAA 1062

196 ProAspLeuIleHisAspLeuLeuGlyHisValProTrpLeuLeuHisPr 212



246 GlnSerAsnLeuIleAlaIleValArgCysPheThrValGlu 262  
1186 GCCACA.....ATTACTGTTACTGTGGAGT 1214  
262 rGlyLeuIleGluAsnHisGluGlyArgLysAlaTyrGlyAlaValLeu 279  
1215 TGGGCTCTCAACAAGAGACTCCATAAGGCAATATGGTGGGCTCC 1264  
279 leSerSerProGlnGluLeuGlyHisAlaPheIleAspAsnValArgVal 295  
1265 TGTCATCCCTTGGTGAATTCAGTACGTCCTTATCAGAGAGCAAGCTT 1314  
296 LeuProGlnGluLeuAspGlnIleAlaArgLeuPropheAsnThrSerTh 312  
1315 CTCCCTGGAGCTGGAGAGACAGCCATCCAAATACACTGTCCACGGA 1364  
312 rProGlnGluThrLeuPheSerIleArgHisPheAspGluLeuValGlu 328  
1365 GTTCCAGCCCCCTGATTATCGTGGCAGAGAGTTTAAATGATGCCAAGGAG 1413

seq name: /SID1/gcgdata/geneseq/geneseqn/NA1999.DAT:AAZ30664

Documentation\_block:

AAZ30664 standard; DNA; 2448 BP.

AC AAZ30664;

15-FEB-2000 (first entry)

Human phenylalanine hydroxylase gene.

Recombinant; hybrid; binding domain; ligand; animal cell; diphtheria;  
translocation domain; botulinum; neurotoxin; ricin; cholera; tetanus;  
shiga-like toxin; pertussis; translocation; cytoplasmic membrane; HIV;  
cytosol; therapy; genetic deficiency disease; enzyme; co-factor; poison;  
adipocyte; cancer; virus; infection; antibody; ss.

Homo sapiens.

US5965406-A.

12-OCT-1999.

07-JUN-1995; 95US-0488246.

04-AUG-1993; 93US-0102387.

07-JUN-1984; 84US-0618199.

27-JUN-1991; 91US-0722484.

25-APR-1985; 85US-0726808.

07-JUN-1985; 85US-0742554.

22-DEC-1989; 89US-0456095.

14-JUN-1990; 90US-0538276.

(SERA-) SERAGEN INC.

Murphy JR;

WPI: 1999-632431/54.

P-PSDB; AAY55893.

Recombinant DNA molecule encoding a three part hybrid protein used in

the treatment of Aids and genetic deficiency diseases -

Example 5; Fig 13; 3lpp; English.

The invention relates to a recombinant DNA molecule encoding a hybrid

protein comprising three parts: (a) the first part comprises a portion

of the binding domain of a cell-binding polypeptide ligand allowing the

hybrid protein to bind to an animal cell; (b) the second part comprises

CC membrane into the cytosol of the cell; and (c) the third part comprises  
CC a polypeptide entity to be introduced into the cell, which is non-native  
CC to the naturally occurring protein of (b). This sequence represents the  
CC human phenylalanine hydroxylase gene for use in generating the hybrid of  
CC the invention. The hybrid molecule enables the direction of appropriate  
CC therapy to affected cells, allowing them to function properly and  
CC alleviate or cure the disease. The hybrid is especially used in treating  
CC genetic deficiency diseases, by delivering to affected cells an enzyme  
CC supplying the missing function, to supplementing cellular levels of a  
CC particular enzyme or a scarce precursor or cofactor, to directing toxins  
CC or other poisons to destroy particular cells (such as adipocytes, cancer  
CC cell, or virus infected-cells), to counteracting viral infections, such as  
CC HIV, by introducing appropriate antibodies to viral proteins. It is also  
CC involved in the process of getting non-therapeutic substances such as  
CC detectable labels into cells.

XX  
SQ Sequence 2448 BP; 739 A; 514 C; 501 G; 694 T; 0 other;

alignment\_scores:

Quality: 213.50 Length: 233

Ratio: 1.642 Gaps: 4

Percent Similarity: 55.794 Percent Identity: 26.609

alignment\_block:

US-09-438-185-1047 x AAZ30664

Align seg 1/1 to: AAZ30664 from: 1 to: 2448

104 ArgAsnLeuTrpTyrArgLeuLeuSerSerArgPheSerLeuTrpLysSe 120

772 ARAAACAATGGGCACAGTGTTCAGACTCTGAGTCTCTGTATATAAC 821

120 rTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeu 137

822 CCATGCT.....TGCTATGAGTACATACATCTTTTCCACCTCTCTG 862

137 erAspPheLeuAspHis.....GlnAlaVal 145

863 AAAAGTACTGTGGCTTCCATGAAGATAACATTCGCCAGCTGGAAGAGT 912

146 IleLysPhePheGluLeuGluThrHisPheSerTyrTyrProValSerG 162

913 TCTCAATTCCTGCAGACTTGCACCTGGTTTCCGCCCTCGACCTGTGGCTG 962

162 yPheValAlaProHisGlnTyrLeuSerLeuGlnAspArgTyrPheP 179

963 CTGCTTTCCTCTCGGGATTCTTGGTGGCTGGCTTCCGAGTCTCTCC 1012

179 rIleAlaSerValMetArgThrLeuAspLysAspAsnPheSerLeuThr 195

1013 ACTGCACACAGTACATCAGACATGGATCCAAAGCCCATGTATACCCCGAA 1062

196 ProAspLeuIleHisAspLeuGlyHisValProThrLeuLeuHisPr 212

1063 CTGACATCTGCATGAGCTGTGGACATGTGCCCTGTGTTTCAGATCG 1112

212 oSerPheSerGluPhePheIleAsnMetGlyArgLeuPheThrLysVal 229

1113 CAGTTTGGCCAGTTTTCAGGAATTTGCC..... 1143

229 leGluLysValGlnAlaLeuProSerLysLysGlnArgIleGlnThrLeu 245

1144 .....CTGCTCTCTGGGTGCACCTGATGATACATATAAGAGCTC 1185

246 GlnSerAsnLeuIleAlaIleValArgCysPheThrPheThrValGluSe 262

1186 GCCACA.....ATTACTGTTTACTGTGGAGT 1214

262 rGlyLeuIleGluAsnHisGluGlyArgLysAlaTyrGlyAlaValLeu 279

1215 TGGGCTCTGCAACAAGAGACTCCATAAGGCAATATGTTGCTGGCTCC 1264

279 leSerSerProGlnGluLeuGlyHisAlaPheIleAspAsnValArgVal 295  
 1265 TGTATCTCTTGGTGAATACAGTACTGCTATACAGAGAGAACCAAGCTT 1314  
 296 LeuProLeuGluLeuAspGlnIleIleArgLeuProPheAsnThrSerTh 312  
 1315 CTCCCTCTGGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1364  
 312 rProGlnGluThrLeuPheSerIleArgHisPheAspGluLeuValGlu 328  
 1365 GTTCCAGCCCTGTATTACGTGGCAGAGAGAGAGAGAGAGAGAGAG 1413  
 seq\_name: /SIDS1/gcgdata/geneseq/NA1991.DAT:AAQ12712  
 seq\_documentation\_block:  
 ID AAQ12712 standard; DNA: 2523 BP.  
 XX AAQ12712;  
 XX 08-OCT-1991 (first entry)  
 XX Phenylalanine hydroxylase.  
 XX Hybrid; fusion; membrane translocation; binding region; HIV;  
 KW infection; toxin; steroid; hormone; monoclonal antibody; antigen;  
 KW diphtheria; exotoxin; phenylketonuria; cholera; interleukin; IL-2;  
 KW protease; epidermal growth factor; ricin; tetanus; hexosaminidase;  
 KW Shiga-like toxin A; SLT-A; PH; ligand; insulin; nuclease; ss.  
 XX Vibrio cholera.  
 XX Key Location/Qualifiers  
 XX CDS 223..1578  
 XX /\*tag= a  
 XX /label= PH  
 XX WO9109871-A.  
 XX 11-JUL-1991.  
 XX 21-DEC-1990; 90WO-US07619.  
 XX 14-JUN-1990; 90US-0538276.  
 XX 22-DEC-1989; 89US-0456095.  
 XX (SERA-) SERAGEN INC.  
 XX Murphy JR;  
 XX WPI: 1991-222845/30.  
 XX P-PSDB; AAR13119.  
 XX Hybrid molecules for targeting chemical entity to cell - have  
 XX membrane trans-locating and cell binding-regions and used to  
 XX treat HIV infection, genetic enzyme-deficiency disorders etc.  
 XX Disclosure; Fig 13(1-3); 59pp; English.  
 XX Hybrid molecules are produced by covalently linking  
 CC (1) a portion (A) of the binding domain of a cell-binding ligand,  
 CC allowing binding of the mol. to an animal cell;  
 CC (2) a portion (B) of a translocation domain of a protein able to  
 CC translocate (C) across the cell cytoplasmic membrane, and  
 CC (3) a portion (C) which is to be introduced into the cell.  
 CC (A) is derived from a steroid or polypeptide hormone, a single-chain  
 CC analogue of a monoclonal antibody able to bind an antigen expressed  
 CC on the cell surface, or a polypeptide toxin.  
 CC (B) is derived from a toxin (e.g. diphtheria toxin or Pseudomonas  
 CC exotoxin A).  
 CC (A) may be derived from insulin, interleukins 2, 3 or 6 or  
 CC epidermal growth factor.  
 CC Suitable enzymes in (C) include cholera toxin, ricin, tetanus toxin,  
 CC hexosaminidase A, protease, nuclease, SLT-A, etc.

CC Specified examples are CT-A/DT-B'/IL-2, SLTA/DT-B'/IL-2,  
 CC ricin A/DT-B'/IL-2, HIV-BP/DT-B' and the phenylalanine  
 CC hydroxylase-DT-B' or their biologically active mutants.  
 CC (CT-A= cholera toxin, DT-B'= truncated diphtheria toxin,  
 CC SLTA= Shiga-like toxin A; HIV-BP= HIV protease binding protein.  
 CC See also AAQ12710-12.  
 XX  
 SQ Sequence 2523 BP; 766 A; 522 C; 518 G; 717 T; 0 other;  
 alignment\_scores:  
 Quality: 213.50 Length: 233  
 Ratio: 1.642 Gaps: 4  
 Percent Similarity: 55.794 Percent Identity: 26.609  
 alignment\_block:  
 US-09-438-185-1047 x AAQ12712  
 Align seg 1/1 to: AAQ12712 from: 1 to: 2523  
 104 ArgAsnLeuTyrPheArgLeuLeuSerSerArgPheSerLeuTrpLysSe 120  
 772 AGAACAATGGGCACAGTGTCAAGACTCTGAAGTCTCTGTATAAAC 821  
 120 rTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeu 137  
 822 CCATGCT.....TGCTATGAGTACATCACAATTTTCCACTCTTG 862  
 137 erAspPheLeuAspHis.....GlnAlaVal 145  
 863 AAAAGTACTGTGGTTCATGAAGATAACATTCGCCAGCTGGAGAGAGCTT 912  
 146 IleLysPhePheGluLeuGluThrHisPheSerTyrTyrProValSerG1 162  
 913 TCTCAATTCCTGCAGACTTGCATGCTGTTCCGCCCTCCGACCTGTGGCTG 962  
 162 yPheValAlaProHisGlnTyrLeuSerLeuLeuGlnAspArgTyrPheP 179  
 963 CTGCTTTCCTCGGATTTCTTGGTGGCTGGCTTCCGAGTCTTCC 1012  
 179 toIleAlaSerValMetArgThrLeuAspLysAspAsnPheSerLeuThr 195  
 1013 ACTGCACACAGTACATCAGACATGGATCCCAAGCCCATGTATACCCCGAA 1062  
 196 ProAspLeuIleHisAspLeuLeuGlyHisValProTyrLeuLeuHisPr 212  
 1063 CCTGACATCTGCATGAGCTGTGGACATGTGCCCTTCTTTTCAGATCG 1112  
 212 oSerPheSerGluPhePheIleAsnMetGlyArgLeuPheThrLysValI 229  
 1113 CAGCTTCCCGAGTTTCCCGAGAAATGCG..... 1143  
 229 leGluLysValGlnAlaLeuProSerLysLysGlnArgIleGlnThrLeu 245  
 1144 .....CTTGCTCTCTGGGTGGACCTGATGATACATTGAAAGCTC 1185  
 246 GlnSerAsnLeuIleAlaIleValArgCysPheTrpPheThrValGluSe 262  
 1186 GCCACA.....ATTACTGTTTACTGTGGAGCTT 1214  
 262 rGlyLeuIleGluAsnHisGluGlyArgLysAlaTyrGlyAlaValLeuI 279  
 1215 TGGGCTCTGCAACAAGAGAGACTCCATAAAGGCATATGCTGGCTGCC 1264  
 279 leSerSerProGlnGluLeuGlyHisAlaPheIleAspAsnValArgVal 295  
 1265 TGTCATCTTTGGTGAATTACAGTACTGCTTATCAGAGAGAGCAAGCTT 1314  
 296 LeuProLeuGluLeuAspGlnIleIleArgLeuProPheAsnThrSerTh 312  
 1315 CTCCCTCTGGAGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1364  
 312 rProGlnGluThrLeuPheSerIleArgHisPheAspGluLeuValGlu 328



seq\_documentation\_block:  
ID AAQ41866 standard; DNA; 1829 BP.

AC AAQ41866;

DT 20-SEP-1993 (first entry)

DE Human tyrosine hydroxylase coding sequence.

KW Human; rat; tyrosine hydroxylase; substitution; N-terminal; dopamine;  
neurological disorder; antibody; variant; ss.

OS Homo sapiens.

FX Key Location/Qualifiers

FT CDS 20..1512

FT /tag= a

XX J05212082-A.

XX PD 18-MAY-1993.

XX PF 13-MAR-1991; 91US-0669446.

XX PR 13-MAR-1991; 91US-0669446.

XX PA (UWNY ) UNIV NEW YORK STATE.

XX PI Filer D, Friedhoff AJ, Goldstein M, Wu J;

XX DR WPI; 1993-175456/21.

XX DR P-PSDB; AAR36740.

XX Genetically modified tyrosine hydroxylase having increased

PT activity - used for treating neurological disorders e.g.

PT Parkinson's and Alzheimer's diseases and affective disorders

XX Disclosure: Fig 6; 20pp; English.

CC This sequence encodes the human tyrosine hydroxylase protein. This  
CC which was used in the construction of a variant tyrosine hydroxylase  
CC which contains at least one amino acid substitution in the N-terminal  
CC 55 amino acids. The substitution corresponds to an amino acid selected  
CC from Ser8, Ser31, Arg37, Arg38, Gln39, Ser40, Leu42, Ile42, Glu43,  
CC Asp44, Ala45, Arg46 or Lys47. The enzymatic activity of the variant  
CC protein is at least 3-fold greater than that for the wild type  
CC protein. Cells transfected with this DNA may be used for treating  
CC neurological disorders associated with a deficiency in tyrosine  
CC hydroxylase or dopamine. The tyrosine hydroxylase protein may be  
CC used to generate antibodies specific for the variant tyrosine  
CC hydroxylases to monitor the enzyme during a treatment regimen.

XX Sequence 1829 BP; 309 A; 636 C; 559 G; 325 T; 0 other;

alignment\_scores:

Quality: 198.00 Length: 219

Ratio: 1.623 Gaps: 4

Percent Similarity: 55.708 Percent Identity: 26.941

alignment\_block:

US-09-438-185-1047 x'AAQ41866 ..

Align seg 1/1 to: AAQ41866 from: 1 to: 1829

128 ASPTYrLeuGluAlaPheGlyLeuLeuSerAspPheLeuAspHis..... 142

766 GAGCACCCTGGAGCGCTCTTCCTTGTGGAGGCTTCAGCGGCTACCGGGA 815

143 .....GlnAlaValIleLysPheGluLeuGluT 153

816 AGACATATATCCCCAGCTGGAGGAGGCTCTCCCGCTTCTGAGGAGCGCA 865

153 hrHisPheSerTyrTyrProValSerGlyPheValAlaProHisGlnTyr 169  
866 GGGGCTTCCAGCTGGCGGCTGTGGCGGCTGCTGTCCCGCGGACTTC 915  
170 LeuSerLeuLeuGlnAspArgTyrPheProIleAlaSerValMetArgTh 186  
916 CTGGCAGCCTGGCGCTTCCCGGCTGTCCAGTCACCCAGTATATCCGCCA 965  
186 rLeuAspLysAspAsnPheSerLeuThrProAspLeuIleHisAspLeuL 203  
966 CGCGTCCVCGCCCATGCACTCCCTGAGCGGAGCTGCTGCCACGAGCTGC 1015  
203 euGlyHisValProTyrPheLeuHisProSerPheSerGluPheHeile 219  
1016 TGGGCGACGTCCCATGCTGTGGCGGACGCACTTCGCGCAGTTCGCGAG 1065  
220 AsnMetGlyArgLeuPheThrLysValIleGluLysValGlnAlaLeuPr 236  
1066 GACATTGGC.....CTGGCGTCCCTGGGGCGCTCGGA 1097  
236 OSerLysLysGlnArgIleGlnThrLeuGlnSerAsnLeuIleAlaIleV 253  
1098 TGAGGAATTCGAGAGCTGTCCACGCTGTCR..... 1128  
253 alArgCysPheThrPheThrValGluSerGlyLeuIleGluAsnHisGlu 269  
1129 .....TGCTTCCAGGTTCGGGCTGTGTAAAGCAGAACGGG 1167  
270 GlyArgLysAlaTyrGlyAlaValLeuIleSerProGlnGlnLeuG1 286  
1168 GAGGTGAAGCCCTATGTGTGGCGGCTGCTCTCTCCTACGGGGAGCTCT 1217  
286 yHisAlaPheIleAspAsnValArgValLeuProLeuGluLeuAspGlnI 303  
1218 GCACCTGCCTGTCTGAGGAGCTGAGATTTCGGGCTTCGACCTGAGGCTG 1267  
303 leIleArgLeuProPheAsnThrSerThrProGlnGlnThrLeuPheSer 319  
1268 CGGCGGTGCAGCCCTACCAAGACGACGAGTACGCTAGCTACTTCGTG 1317  
320 IleArgHisPhe.....AspGluLeuValGluLeuThrSerLy 332  
1318 TCTGAGAGCTTCAGTCAGCGCCAGGACACAGCTCAGGAGCTATCCCTCAGG 1367  
332 sLeuGlu 334  
1368 CATCCAG 1374

seq\_name: /STD1/gcgdata/geneseq/geneseq/NA2000.DAT:AAA91625

seq\_documentation\_block:

ID AAA91625 standard; cDNA; 817 BP.

XX AAA91625;

XX 12-JAN-2001 (first entry)

XX Caenorhabditis elegans cod-5 knockout cDNA.

XX Caenorhabditis elegans; cod-5; daf-18; insulin signalling pathway;

XX daf-2; age-1; insulin receptor; PI 3-kinase; PKB kinase;

XX PTEN lipid phosphatase; antidiabetic; anorectic; obesity; diabetes; ss.

XX Caenorhabditis elegans.

XX Synthetic.

XX Key Location/Qualifiers

XX CDS 40..177

XX /tag= a

XX /product= "Cod-5 frameshift mutant"

XX WO200033068-A1.





Mon Jul 2 08:46:42 2001

us-09-438-185-1047.rng

Page 16

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2001, 13:51:51 ; Search time 31 Seconds  
(without alignments)  
400.016 Million cell updates/sec

Title: US-09-438-185-1047

Perfect score: 1889

Sequence: 1 VHCERTLPKYLKIAKL.....ESIPLYNQEKYLSGFVQLQ 362

Scoring table: BLOSUM62  
Gap 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_39.:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	235.5	12.5	262	1 PH4H_PSEAE	P43334 pseudomonas
2	231	12.2	445	1 TR5H_CHICK	P70080 gallus gall
3	230.5	12.2	444	1 TR5H_HUMAN	P17752 homo sapien
4	229.5	12.1	491	1 TY3H_PHASP	P11982 phasianidae
5	227.5	12.0	444	1 TR5H_RAT	P09810 rattus norv
6	226.5	12.0	447	1 TR5H_MOUSE	P17532 mus musculus
7	225.5	11.9	444	1 TR5H_RABIT	P17290 oryctolagus
8	225.5	11.9	488	1 TY3H_ANGAN	O42091 anguilla an
9	225.5	11.9	498	1 TY3H_MOUSE	P24529 mus musculus
10	223.5	11.8	481	1 TR5H_XENLA	Q92142 xenopus lae
11	223	11.8	490	1 TY3H_RAT	P04177 rattus norv
12	221.5	11.7	490	1 TY3H_BOVIN	P17289 bos taurus
13	220	11.6	528	1 TY3H_HUMAN	P07101 homo sapien
14	219	11.6	508	1 TY3H_DROME	P18459 drosophila
15	216.5	11.5	465	1 TY3H_SCHMA	O17446 schistosoma
16	214	11.3	453	1 PH4H_RAT	P04176 rattus norv
17	213.5	11.3	452	1 PH4H_HUMAN	P00439 homo sapien
18	212.5	11.2	453	1 PH4H_MOUSE	P16331 mus musculus
19	194	10.3	452	1 PH4H_DROME	P17276 drosophila
20	190	10.1	281	1 PH4H_CHRVO	P30967 chromobacte
21	178.5	9.4	404	1 TY3H_CABEL	P30986 caenorhabdi
22	175.5	9.3	457	1 PH4H_CABEL	P30925 caenorhabdi
23	166.5	5.6	1501	1 NINL_DROME	P10676 drosophila
24	164.5	5.5	2331	1 RYSL_MABVP	P35262 marburg vir
25	162.5	5.4	563	1 YR5H_SCHPO	Q10144 schizosacch
26	160.5	5.3	1157	1 N133_YEAST	P36161 saccharomyc
27	160.5	5.3	2212	1 RRPL_EBOZM	Q05318 ebola virus
28	98.5	5.2	1024	1 Y075_MYCGE	P47321 mycoplasma
29	97	5.1	492	1 CPL1_PANAR	O27712 panulirus a
30	97	5.1	1050	1 EX5B_CHLPN	Q92767 chlamydia p
31	96.5	5.1	512	1 G6PD_CHLPN	Q92806 chlamydia p
32	96	5.1	529	1 YABD_SCHPO	Q09812 schizosacch
33	96	5.1	2059	1 TEGU_HSV7J	P52362 human herpe

34	95.5	5.1	1398	1 TOP2_PLAPK	P41001 plasmodium
35	95	5.0	1427	1 SRB8_YEAST	P25648 saccharomyc
36	95	5.0	1475	1 TRA2_CABEL	P34709 caenorhabdi
37	93.5	4.9	650	1 Y096_MYCGE	P47342 mycoplasma
38	92.5	4.9	466	1 Y065_MYCGE	P47311 mycoplasma
39	92	4.9	447	1 FLP_XLULA	P13783 kluyveromyc
40	92	4.9	526	1 MATK_ALUAIH	P56784 tritidopsis
41	92	4.9	544	1 MATK_MAIZE	P48190 zea mays (m
42	92	4.9	4563	1 APB_HUMAN	P04114 homo sapien
43	91.5	4.8	293	1 SPEE_METJA	Q57761 mehanococc
44	91	4.8	2183	1 RRPL_MEASA	P35975 measles vir
45	91	4.8	2183	1 RRPL_MEASE	P12576 measles vir

#### ALIGNMENTS

RESULT 1  
PH4H\_PSEAE  
ID PH4H\_PSEAE STANDARD; PRT: 262 AA.  
AC P43334;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE PHENYLALANINE-4-HYDROXYLASE (EC 1.14.16.1) (PAH) (PHE-4-  
DE MONOOXYGENASE).  
GN PHA OR PA0872.  
OS Pseudomonas aeruginosa.  
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
OC Pseudomonas.  
OX NCBI\_TaxID=287;  
[...]  
RN SEQUENCE FROM N.A.  
RP STRAIN=ATCC 15692 / PA01;  
RC MEDLINE=94151331; PubMed=8108417;  
RX Zhao G., Xia T., Song J., Roy R.A.;  
RT "Pseudomonas aeruginosa possesses homologues of mammalian  
RT phenylalanine hydroxylase and 4 alpha-carbinolamine dehydratase/DCoH  
RT as part of a three-component gene cluster."  
RL Proc. Natl. Acad. Sci. U.S.A. 91:1366-1370(1994).  
[2]  
RN SEQUENCE FROM N.A.  
RP STRAIN=PA01;  
RC MEDLINE=20437337; PubMed=10984043;  
RX Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,  
Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,  
Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,  
Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,  
Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;  
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
RT opportunistic pathogen."  
RT Nature 406:959-964(2000).  
RL  
CC -1- CATALYTIC ACTIVITY: L-PHENYLALANINE + TETRAHYDROBIOTERIN + O(2) -  
CC -1- L-TYROSINE + DIHYDROBIOTERIN + H(2)O.  
CC -1- COFACTOR: FERROUS ION.  
CC -1- PATHWAY: CATABOLISM OF PHENYLALANINE; FIRST, RATE-LIMITING STEP.  
CC -1- SIMILARITY: MONOMER.  
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID  
CC HYDROXYLASES FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL: M88627; AAA25936.1;  
CC EMBL: AB004522; BAG04261.1;  
CC HSSP: P04177; 1TQH  
CC InterPro: IPR001273;

DR Pfam; PF00351; biotin.H; 1.  
DR PROSITE; PS00367; BIOTERIN\_HYDROXYL; 1.  
KW Oxidoreductase; Monooxygenase; Phenylalanine catabolism; Iron.  
FT METAL 121 121 IRON (BY SIMILARITY).  
FT METAL 126 126 IRON (BY SIMILARITY).  
FT CONFLICT 135 135 F -> L (IN REF. 1).  
SQ SEQUENCE 262 AA; 30322 MW; A565839C5961A45 CRC64;  
  
Query Match 12.5%; Score 235.5; DB 1; Length 262;  
Best Local Similarity 24.3%; Pred. No. 9e-12;  
Matches 61; Conservative 52; Mismatches 103; Indels 33; Gaps 6;  
QY 93 DNAYVSTFFNNRWLLSRLLSRLLSKYCPREFLDYLEAFGLLDFDHOAVKFELE 152  
DB 13 DNGFHYPTHQVNTLITQLKVIKGRACQYLGIEQLG-----LPHKIFOLDEIN 67  
QY 153 -----THFSYVPSGVFVAPHQYLSLLQDRYPIASVMTLDKDNFSLTPDLIHDLLGHV 206  
DB 68 RVLQATQGWVRVAPALIPFQTFELLASQFFVATFIRPELDYLGQPDIFHEIFGHC 127  
QY 207 PWWLHPSEFFNMGRLFTKVIKQVQALPKKRIQTLQSLNLAIVRCFWFTVESGLIE 266  
DB 128 PLITNPFAEFTYTGKLGKA-----SKER-----VFLARLYNMTIEFGLVE 171  
QY 267 NHEGRKAYGAVLISPOLGHAFTD--NVRVLPLELDQIIRLPENTSTPQETLFSIRHF 323  
DB 172 TDQGRKRYGGILSPKTVYSLSDEPLQAFNPLE---AMRTPYRIDILQPLVFLPDL 228  
QY 324 DELVELTSK 332  
DB 229 KRLFOLAQE 237  
  
RESULT 2  
ID TR5L CHICK STANDARD; PRT; 445 AA.  
AC P70860;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE TRYPTOPHAN 5-MONOOXYGENASE (EC 1.14.16.4) (TRYPTOPHAN 5-HYDROXYLASE).  
GN TPH.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
CC Gallus.  
CC NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-WHITE LEGHORN; TISSUE-Pineal gland;  
RX MEDLINE=97072811; PubMed=8915576;  
RA Florez J.C., Seidenman K.J., Barrett R.K., Sangoram A.M.,  
RA Takahashi J.S.;  
RT "Molecular cloning of chick pineal tryptophan hydroxylase and circadian oscillation of its mRNA levels";  
RL Brain Res. Mol. Brain Res. 42:25-30(1996).  
CC -1- CATALYTIC ACTIVITY: L-TRYPTOPHAN + TETRAHYDROPTERIDINE + O(2) =  
CC 5-HYDROXY-L-TRYPTOPHAN + DIHYDROPTERIDINE + H(2)O.  
CC -1- COFACTOR: FERROUS ION.  
CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEAL GLAND.  
CC -1- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID HYDROXYLASES FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).  
CC EMBL; U26428; AAC60036.1; -  
DR HSP; P04177; ITOH.  
DR InterPro; IPR001273; -  
DR InterPro; IPR002912; -  
DR Pfam; PF01842; ACT; 1.  
DR Pfam; PF00351; biotin.H; 1.  
DR PRINTS; PR00372; FYWHYDRXLA.  
DR PROSITE; PS00367; BIOTERIN\_HYDROXYL; 1.  
KW Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;  
KW phosphorylation. 58 PHOSPHORYLATION (BY CAPE) (POTENTIAL).  
FT METAL 273 273 IRON (BY SIMILARITY).  
FT METAL 278 278 IRON (BY SIMILARITY).  
FT METAL 318 318 IRON (BY SIMILARITY).  
SQ SEQUENCE 445 AA; 51139 MW; FF0041D7C4B159F6 CRC64;  
  
Query Match 12.2%; Score 231; DB 1; Length 445;  
Best Local Similarity 28.4%; Pred. No. 4.1e-11;  
Matches 65; Conservative 41; Mismatches 93; Indels 30; Gaps 4;  
QY 108 YRLSSRFSLKSYCPREFLDYLEAFGLLSDFLDH-----QAVIKFELETHESYYP 159  
DB 179 YRELN-----KLYPHACREYLNKLPILTKYCYREDNIPQLEDVSRFLKERTGET 232  
QY 160 VSGFVAPHQYLSLLQDRYPIASVMTLDKDNFSLTPDLIHDLLGHVPLHPSFSEFFI 219  
DB 233 VAGYLSPRDFLAGLAFRVFCHTQYVRHSSDPLYTPEDTCHHELLGHVPLLAEPFSAQFSQ 292  
QY 220 NMGRLETKVIEKQVQALPKKRIQTLQSLNLAIVRCFWFTVESGLIENHGRKAYGAVLI 279  
DB 293 EIG-----LASLGASDEAVOKLAT-----CYFTVEFGCKQEGOLRVYGALL 336  
QY 280 SSPQELGHAFIDNVRVLPLELDQIIRLPENTSTPQETLFSIRHDELVE 328  
DB 337 SSISELAHSLSGAKVPPDPKVTCKQECCLITTFQEVTFVSEFEAKE 385  
  
RESULT 3  
ID TR5H HUMAN STANDARD; PRT; 444 AA.  
AC P17752; Q16736;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE TRYPTOPHAN 5-MONOOXYGENASE (EC 1.14.16.4) (TRYPTOPHAN 5-HYDROXYLASE).  
GN TPH OR TRPH.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
CC NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Carcinoma;  
RX MEDLINE=90332431; PubMed=2377472;  
RA Boulard S., Barmon M.C., Ganem Y., Launay J.M., Mallet J.;  
RT "Complete coding sequence of human tryptophan hydroxylase";  
RL Nucleic Acids Res. 18:4257-4257(1990).  
CC [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE-Brain;  
RX MEDLINE=95077422; PubMed=7986090;  
RT Tipper J.P., Citron B.A., Ribeiro P., Kaufman S.;  
RT "Cloning and expression of rabbit and human brain tryptophan hydroxylase cDNA in Escherichia coli";  
RL Arch. Biochem. Biophys. 315:445-453(1994).  
CC -1- CATALYTIC ACTIVITY: L-TRYPTOPHAN + TETRAHYDROPTERIDINE + O(2) =  
CC 5-HYDROXY-L-TRYPTOPHAN + DIHYDROPTERIDINE + H(2)O.  
CC -1- COFACTOR: FERROUS ION.  
CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST

CC STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEL GLAND.  
CC -1- SUBUNIT: MOLTIMER OF IDENTICAL SUBUNITS.  
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID  
CC HYDROXYLASES FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X52836; CAA37018.1; -;  
DR EMBL; L29306; AAA67050.1; -;  
DR PIR; S10489; S10489.  
DR HSP; P04177; ITOH.  
DR MIM; 191060; -;  
DR InterPro; IP001273; -;  
DR InterPro; IP002912; -;  
DR Pfam; PF01842; ACT.1.  
DR PRINTS; PR00351; bioprotein.H.1.  
DR PRINTS; PR00372; FWHYDEXASE.  
DR PROSITE; PS00367; BIOTERIN-HYDROXYL.1.  
KW Oxidoreductase; Monooxygenase; Seroconin biosynthesis; Iron;  
KW Phosphorylation.  
FT MOD\_RES 58 58 PHOSPHORYLATION (BY CAPK) (POTENTIAL).  
FT METAL 272 272 IRON (BY SIMILARITY).  
FT METAL 277 277 IRON (BY SIMILARITY).  
FT METAL 317 317 IRON (BY SIMILARITY).  
FT CONFLICT 19 19 T -> S (IN REF. 1).  
FT CONFLICT 68 68 T -> I (IN REF. 1).  
FT CONFLICT 90 91 TP -> NL (IN REF. 1).  
FT CONFLICT 97 97 M -> L (IN REF. 1).  
FT CONFLICT 100 100 E -> D (IN REF. 1).  
FT CONFLICT 104 104 S -> T (IN REF. 1).  
FT CONFLICT 151 151 L -> S (IN REF. 2).  
FT CONFLICT 154 154 N -> S (IN REF. 2).  
FT CONFLICT 157 157 H -> Y (IN REF. 2).  
FT CONFLICT 179 179 R -> Q (IN REF. 1).  
FT CONFLICT 207 207 R -> Q (IN REF. 2).  
FT CONFLICT 217 217 V -> I (IN REF. 2).  
FT CONFLICT 344 344 A -> V (IN REF. 2).  
FT CONFLICT 414 414 T -> A (IN REF. 2).  
FT CONFLICT 419 419 S -> N (IN REF. 2).  
FT CONFLICT 425 425 R -> Q (IN REF. 1).  
FT CONFLICT 436 436 A -> G (IN REF. 2).  
SQ SEQUENCE 444 AA; 51032 MW; 86C39884F72E120A CRC64;  
Query Match 12.2%; Score 230.5; DB 1; Length 444;  
Best Local Similarity 27.8%; Pred. No. 4.5e-11;  
Matches 64; Conservative 43; Mismatches 96; Indels 27; Gaps 4;  
QY 107 WYRLSSRSLKWSYCPREFLDYLEAFGLSLDFLDH-----QAVIKFFETHFSY 158  
Db 174 WGTFRNLKLYTHACR---EYLKNEPLLSKYGYREDNIPQLEDVSNFLKERTGFSIR 230  
QY 159 PVSGFVAPHOYLSLLQDRYFPPIASVMRTLDKDNFSLPDLIHLGLHVPWLLHPSPSEFF 218  
Db 231 PVAGYLSRFLSLGLAFRVFCHCTQYVRHSDPFTPEPTDCHELLGHVPLLAEPSSAQFS 290  
QY 219 INMGRLFTKVIKVOALPSKKORIOTLQSLNLAIVRCFWFTVSGLIENHEGRKAYGAVL 278  
Db 291 QEIG-----LASIGASEAVQKAT-----CYFTVEFLGKODGQLRVGAGL 334  
QY 279 ISSPQELGHAFIDNVRVLPLELDQIIRLPFNTPSTPQETLSIRHFDE 325  
Db 335 LSSISELKHLSCHAKVFFDKTKCKECLITTFQDVYVFSSEFDEAKE 384

RESULT 4  
TY3H\_PHASP

ID TY3H\_PHASP STANDARD; PRT; 491 AA.  
AC P11982; 1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 01-OCT-1989 (Rel. 35, Last annotation update)  
DE TYROSINE 3-MONOXYGENASE (EC 1.14.16.2) (TYROSINE 3-HYDROXYLASE) (TH).  
GN TH.  
OS Phasianidae sp. (Quail).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae.  
OX NCBI\_TaxID=9006;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Adrenal gland;  
RX MEDLINE=88089590; PubMed=2447231;  
RA Fauquet M., Grima B., Lamouroux A., Mallet J.;  
RT "Cloning of quail tyrosine hydroxylase: amino acid homology with  
other hydroxylases discloses functional domains.";  
RL J. Neurochem. 50:142-148(1988).  
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC  
CC NEURONES.  
CC -1- CATALYTIC ACTIVITY: L-TYROSINE + TETRAHYDROPTERIDINE + O(2) =  
CC 3,4-DIHYDROXY-L-PHENYALANINE + DIHYDROPTERIDINE + H(2)O.  
CC -1- COFACTOR: FERROUS ION.  
CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE  
CC CATALYTIC ACTIVITY.  
CC -1- PATHWAY: RATE-LIMITING STEP IN CATECHOLAMINE BIOSYNTHESIS.  
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID  
CC HYDROXYLASES FAMILY.  
CC -----  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; M24778; AAA49514.1; -;  
DR PIR; A28582; A28582.  
DR HSP; P04177; ITOH.  
DR InterPro; IP001273; -;  
DR Pfam; PF00351; bioprotein.H.1.  
DR PRINTS; PR00372; FWHYDEXASE.  
DR PROSITE; PS00367; BIOTERIN-HYDROXYL.1.  
DR Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;  
KW Neurotransmitter biosynthesis; Phosphorylation.  
KW MOD\_RES 40 40 PHOSPHORYLATION (BY CAPK)  
FT DOMAIN 51 55 POLY-ALA.  
FT METAL 324 324 IRON (BY SIMILARITY).  
FT METAL 329 329 IRON (BY SIMILARITY).  
FT METAL 369 369 IRON (BY SIMILARITY).  
SQ SEQUENCE 491 AA; 56066 MW; AFB363220F70C0A0 CRC64;  
Query Match 12.1%; Score 229.5; DB 1; Length 491;  
Best Local Similarity 29.1%; Pred. No. 6.1e-11;  
Matches 66; Conservative 34; Mismatches 100; Indels 27; Gaps 4;  
QY 107 WYRLSSRSLKWSYCPREFLDYLEAFGLSLDFLDH-----QAVIKFFETHFSY 158  
Db 236 WKEYSTLSLYPTHACK---EYLEAFNLLEKFCGYNNNTIPQLEFVSFLKERTGQFLR 282  
QY 159 PVSGFVAPHOYLSLLQDRYFPPIASVMRTLDKDNFSLPDLIHLGLHVPWLLHPSPSEFF 218  
Db 283 PVRLGLSARDFLASLAFRVFQCTQYIRHASSPMHSPEDCCHELLGHVPLMDKTFQAQS 342  
QY 219 INMGRLFTKVIKVOALPSKKORIOTLQSLNLAIVRCFWFTVSGLIENHEGRKAYGAVL 278  
Db 343 QDIG-----LASIGATDEIEKATL-----YWFVEGLCRQNGIYKAGACL 386  
QY 279 ISSPQELGHAFIDNVRVLPLELDQIIRLPFNTPSTPQETLSIRHFDE 325

Db 387 LSSYGEIHLISLDEPVRDFDPAADAAVOPCQDPQPVYFVSEFSD 433

```
RESULT 5
TRSH_RAT TRSH_RAT STANDARD; PRT; 444 AA.
AC P09810;
DT 01-WAR-1989 (Rel. 10, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TRYPTOPHAN 5-MONOOXYGENASE (EC 1.14.16.4) (TRYPTOPHAN 5-HYDROXYLASE).
GN TPH.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RE SEQUENCE FROM N.A.
RX STRAIN-WISTAR; TISSUE=Pineal gland;
RX MEDLINE=88244702; PubMed=3379411;
RA Darmon M.C., Guilbert B., Levieil V., Ehret M., Maitre M., Mallet J.;
RT "Sequence of two mRNAs encoding active rat tryptophan hydroxylase.";
RL J. Neurochem. 51:312-316(1988).
RN [2]
RE SEQUENCE FROM N.A.
RX MEDLINE=91245924; PubMed=1645430;
RA Kim K.S., Wessel T.C., Stone D.M., Carver C.H., Joh T.H., Park D.H.;
RT "Molecular cloning and characterization of cDNA encoding tryptophan
hydroxylase from rat central serotonergic neurons.";
RL Brain Res. Mol. Brain Res. 9:277-283(1991).
RN [3]
RE SEQUENCE OF 167-261 FROM N.A.
RX MEDLINE=87005247; PubMed=2875901;
RA Darmon M.C., Grima B., Cash C.D., Maitre M., Mallet J.;
RT "Isolation of a rat pineal gland cDNA clone homologous to tyrosine
and phenylalanine hydroxylases.";
RL FEBS Lett. 206:43-46(1986).
CC -1- CATALYTIC ACTIVITY: L-TRYPTOPHAN + TETRAHYDROPTERIDINE + O(2) =
5-HYDROXY-L-TRYPTOPHAN + DIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF
SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST
STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEAL GLAND.
CC -1- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
HYDROXYLASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M28000; AAA2262.1;
DR EMBL; X53501; CAA37579.1;
DR PIR; J10034; WHRTW.
DR PIR; A24367; A24367.
DR PIR; A60034; A60034.
DR HSP; P04177; ITOH.
DR InterPro; IPR001273;
DR InterPro; IPR002912;
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00351; bioplerin_1.
DR PRINTS; PR00372; FYWHYDRXLASE.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
KW Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;
KW Phosphorylation.
FT MOD_RES 58 58 PHOSPHORYLATION (BY CAPK) (POTENTIAL).
FT METAL 272 272 IRON (BY SIMILARITY).
FT METAL 277 277 IRON (BY SIMILARITY).
```

```
FT METAL 317 317 IRON (BY SIMILARITY).
SQ SEQUENCE 444 AA; 51068 MW; C3CF5245727CC825 CRC64;

Query Match 12.0%; Score 227.5; DB 1; Length 444;
Best Local Similarity 27.4%; Pred. No. 7.7e-11;
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;

QY 107 WYRLSSRFSLKSYCFRFLDYLEAFGLSLDFLDH-----QAVIKPELETHFSY 158
DB 174 WGTFRRLNKLYTHACR---EYLRNPLSKYCGYREDNVPOLEDVSNELKERTGFSIR 230
QY 159 PVSGFVAPHOYLSLQDRPPTASVMTLDKDNFSLTDLHLLGHVPMLLPSPSEFF 218
DB 231 PVAGYLSRPFSLGLAFRVFHCQYVRRHSDDPLYTPEDTCHELLGHVPLLAEPSPAQFS 290
QY 219 INMGRLFTKVIKRVQALPSKKRIQTLQSLNLAIVRCFWFTVESGLIENHGRKAYGAVL 278
DB 291 QEIG-----LASLGASEETVOKLAT-----CYFTVFGLCKQDQURVFGAGL 334
QY 279 ISSPQELGHAFIDNVRYVLPLELDQIIRLPFNFTSTPQETLFSIRHFDLVE 328
DB 335 LSSISELHSHALSHGAKVPDPKACKQELITSFQDVYFVSEFDEAKE 384

RESULT 6
TRSH_MOUSE TRSH_MOUSE STANDARD; PRT; 447 AA.
AC P17532;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TRYPTOPHAN 5-MONOOXYGENASE (EC 1.14.16.4) (TRYPTOPHAN 5-HYDROXYLASE).
GN TPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RE SEQUENCE FROM N.A.
RX MEDLINE=90243261; PubMed=2110547;
RA Stoll J., Kozak C.A., Goldman D.;
RT "Characterization and chromosomal mapping of a cDNA encoding
tryptophan hydroxylase from a mouse mastocytoma cell line.";
RL Genomics 7:88-96(1990).
CC -1- CATALYTIC ACTIVITY: L-TRYPTOPHAN + TETRAHYDROPTERIDINE + O(2) =
5-HYDROXY-L-TRYPTOPHAN + DIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF
SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST
STEP OF THE SYNTHESIS OF MELATONIN IN THE PINEAL GLAND.
CC -1- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS.
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID
HYDROXYLASES FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04758; AAA63401.1;
DR PIR; A34582; A34582.
DR HSP; P04177; ITOH.
DR MGD; MGI:98796; Tph.
DR InterPro; IPR001273;
DR InterPro; IPR002912;
DR Pfam; PF01842; ACT; 1.
DR Pfam; PF00351; bioplerin_H; 1.
DR PRINTS; PR00372; FYWHYDRXLASE.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
KW PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
```

KW Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;  
KW MOD\_RES 61 PHOSPHORYLATION (BY CAPK) (POTENTIAL).  
FT METAL 275 IRON (BY SIMILARITY).  
FT METAL 280 IRON (BY SIMILARITY).  
FT METAL 320 IRON (BY SIMILARITY).  
SQ SEQUENCE 447 AA; 51343 MW; 16C839F22A138BCA CRC64;  
  
Query Match 12.0%; Score 226.5; DB 1; Length 447;  
Best Local Similarity 27.0%; Pred. No. 9.3e-11;  
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;  
  
QY 107 WYRLSSRFLSKSYCPREFDYLEAFGLLSDFLDH-----QAVIKFFELETHFSY 158  
Db 177 WGTIFRELKLYPTHACR---EYLRLPLLSKYCYREDNIPQLEDVSNFLKERTGSIR 233  
QY 159 PVSGFVAPHOYLSDQDRYFPIASVMTLQDNFSLTPDLIHLGHVPLWLLHPSFSEFF 218  
Db 234 PVAGYLSRDFLSGLAFVHCTQYVHSSDPLTPEDTCHELLGHVPLLAEPFSAQFS 293  
219 INNGRLFTVKIEKVALPSKQRIOTLOSNIAlVRCFWFTVESGLIENHGRKAYGAVL 278  
Db 294 QEIG-----LASLGASEETVOKLAT-----CYFTVEFGCKODGQQLRVFGAGL 337  
QY 279 ISSPQELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHFDLVE 328  
Db 338 LSSISELKHVLGHAKVPDPKITYKOECLITTFQDVYFVSFEDAKE 384  
  
RESULT 7  
TR5H\_RABIT STANDARD; PRT: 444 AA.  
AC P17290; Q29522;  
DT 01-AUG-1990 (Rel. 15, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE TRYPTOPHAN 5-MONOOXYGENASE (EC 1.14.16.4) (TRYPTOPHAN 5-HYDROXYLASE).  
GN TPH.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=87289638; PubMed=3473690;  
RA Grenett H.E., Ledley F.D., Reed L.L., Woo S.L.C.;  
FT "Full-length cDNA for rabbit tryptophan hydroxylase: functional  
domains and evolution of aromatic amino acid hydroxylases.";  
Proc. Natl. Acad. Sci. U.S.A. 84:5530-5534(1987).  
[2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RX MEDLINE=95077422; PubMed=7986090;  
RA Tipper J.P., Citron B.A., Ribeiro P., Kaufman S.;  
FT "Cloning and expression of rabbit and human brain tryptophan  
hydroxylase cDNA in Escherichia coli.";  
Arch. Biochem. Biophys. 315:445-453(1994).  
RL Arch. Biochem. Biophys. 315:445-453(1994).  
CC -1- CATALYTIC ACTIVITY: L-TRYPTOPHAN + TETRAHYDROPTERIDINE + H(2)O.  
CC -1- COFACTOR: FERROUS ION.  
CC -1- PATHWAY: THIS IS THE RATE-LIMITING ENZYME IN THE BIOSYNTHESIS OF  
SEROTONIN IN THE CENTRAL NERVOUS SYSTEM AND CATALYZES THE FIRST  
STEP OF THE SYNTHESIS OF MELATONIN IN THE PINAL GLAND.  
CC -1- SUBUNIT: MULTIMER OF IDENTICAL SUBUNITS.  
CC -1- SIMILARITY: BELONGS TO THE BIOTRAN-DEPENDENT AROMATIC AMINO ACID  
HYDROXYLASES FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M17250; AAA31487.1; -;  
DR EMBL; L29305; AAA67051.1; -;  
DR PIR; A32699; A32699.  
DR HSSP; P04177; ITOH.  
DR InterPro; IPR001273; -;  
DR InterPro; IPR002912; -;  
DR Pfam; PF01842; ACT; 1.  
DR Pfam; PF00351; bioperlin\_H; 1.  
DR PRINTS; PR00372; FYNHDXLASE.  
DR PROSITE; PS00367; BIOTRAN\_HYDROXYL; 1.  
DR Oxidoreductase; Monooxygenase; Serotonin biosynthesis; Iron;  
KW Phosphorylation.  
KW MOD\_RES 58 58 PHOSPHORYLATION (BY CAPK) (POTENTIAL).  
FT METAL 272 272 IRON (BY SIMILARITY).  
FT METAL 277 277 IRON (BY SIMILARITY).  
FT METAL 317 317 IRON (BY SIMILARITY).  
FT METAL 317 317 IRON (BY SIMILARITY).  
FT CONFLICT 102 102 M -> L (IN REF. 1).  
FT CONFLICT 151 151 L -> S (IN REF. 2).  
FT CONFLICT 151 151 KY -> ND (IN REF. 1).  
FT CONFLICT 202 203 R -> Q (IN REF. 2).  
FT CONFLICT 207 207 T -> K (IN REF. 1).  
FT CONFLICT 390 390  
SQ SEQUENCE 444 AA; 51118 MW; BF182451B2BEC80 CRC64;  
  
Query Match 11.9%; Score 225.5; DB 1; Length 444;  
Best Local Similarity 27.0%; Pred. No. 1.1e-10;  
Matches 62; Conservative 44; Mismatches 97; Indels 27; Gaps 4;  
  
QY 107 WYRLSSRFLSKSYCPREFDYLEAFGLLSDFLDH-----QAVIKFFELETHFSY 158  
Db 174 WGTIFRELKLYPTHACR---EYLKPLLSKYCYREDNIPQLEDVSNFLKERTGSIR 230  
QY 159 PVSGFVAPHOYLSDQDRYFPIASVMTLQDNFSLTPDLIHLGHVPLWLLHPSFSEFF 218  
Db 231 PVAGYLSRDFLSGLAFVHCTQYVHSSDPLTPEDTCHELLGHVPLLAEPFSAQFS 290  
QY 219 INNGRLFTVKIEKVALPSKQRIOTLOSNIAlVRCFWFTVESGLIENHGRKAYGAVL 278  
Db 291 QEIG-----LASLGASEAVOKLAT-----CYFTVEFGCKODGQQLRVFGAGL 334  
QY 279 ISSPQELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHFDLVE 328  
Db 335 LSSISELKHVLGHAKVPDPKITYKOECLITTFQDVYFVSFEDAKE 384  
  
RESULT 8  
TY3H\_ANGAN STANDARD; PRT: 488 AA.  
ID TY3H\_ANGAN  
AC O42091;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE TYROSINE 3-MONOOXYGENASE (EC 1.14.16.2) (TYROSINE 3-HYDROXYLASE) (TH).  
GN TH.  
OS Anguilla anguilla (European freshwater eel).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Anguilliformes; Anguillidae;  
OC Anguillidae; Anguilla.  
OX NCBI\_TaxID=7936;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=9834760; PubMed=9681435;  
RA Bouliard S., Biguet N.F., Vidal B., Veron M., Mallet J.,  
RT Vincent J.D., Dufour S., Vernier P.;  
FT "Tyrosine hydroxylase in the European eel (Anguilla anguilla): cDNA  
cloning, brain distribution, and phylogenetic analysis.";  
J. Neurochem. 71:460-470(1998).  
RL J. Neurochem. 71:460-470(1998).  
CC -1- FUNCTION: PLAYS AN IMPORTANT  
CC NEURONES (BY SIMILARITY).  
CC -1- CATALYTIC ACTIVITY: L-TYROSINE + TETRAHYDROPTERIDINE + O(2) =

```

CC 3,4-DIHYDROXY-L-PHENYLALANINE + DIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE
CC CATALYTIC ACTIVITY (BY SIMILARITY).
CC -1- PATHWAY: RATE-LIMITING STEP IN CATECHOLAMINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AJ000731; CAA04264.1; -
CC HSSP: P04177; ITOH.
CC InterPro: IPR001273; -
CC Pfam: PF00351; biopterin_H; 1.
CC PRINTS: PR00372; FWHYDRXLASE.
CC PROSITE: PS00367; BIOPTERIN_HYDROXYL; 1.
CC Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
CC Neurotransmitter biosynthesis; Phosphorylation.
CC MOD_RES 38 38 PHOSPHORYLATION (BY CAPK) (BY
CC SIMILARITY).
CC METAL 321 321 IRON (BY SIMILARITY).
CC METAL 326 326 IRON (BY SIMILARITY).
CC METAL 366 366 IRON (BY SIMILARITY).
CC SEQUENCE 488 AA; 55490 MW; 573553BA39EBD448 CRC64;
CC -----
Query Match 11.9%; Score 225.5; DB 1; Length 488;
Best Local Similarity 28.1%; Pred. No. 1.3e-10;
Matches 65; Conservative 38; Mismatches 99; Indels 29; Gaps 5;
QY 107 WYRLSSRFLSKSY-CPRFFLDYLEAFGLSLDFLDH-----QAVIKFELETHFSY 157
DB 223 WREYVSTLRLDYTHACS---EHLEAFLLERHCYSPNSIPQLEDVSHLKERTGFQL 278
QY 158 YVSGFVAPHOYLSLQDQRYFPPIASVMRTLDKDNFSLTDLIHLGLHVPMLLPSPSEFF 217
DB 279 RPVAGLLSARDFLASLAFRVFOCTQYIRHASSPMHSPEDPCVHELLGHVPLMDRTTAQFS 338
QY 218 FNMGRLETKVIEKQVQALPSKQRIQTLQSNLIAIVRCFWFTVESGLIENHEGRKAYGAV 277
339 SONIG-----LASLGASDEIEKLST-----LWFTVEFGCKQGDGVKAYGAG 382
QY 278 LSSPQELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHDELVE 328
DB 383 LSSYGLVHLSDEPERREFDPEAAAEPYQDNQYQVYFVSESFTDAKE 433
RESULT 9
ID TY3H_MOUSE STANDARD; PRT: 498 AA.
AC P24529;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TYROSINE 3-MONOOXYGENASE (EC 1.14.16.2) (TYROSINE 3-HYDROXYLASE) (TH).
GN TH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91248263; PubMed=1674869;
RA Ichikawa S., Sasaoka T., Nagatsu T.;
RT "Primary structure of mouse tyrosine hydroxylase deduced from its
RT cDNA.";
RL Biochem. Biophys. Res. Commun. 176:1610-1616(1991).

```

```

RN SEQUENCE OF 1-30 FROM N.A.
RP STRAIN=BALB/C;
RA Morgan W.W., Bermudez J., Sharp Z.D.;
RL Submitted (APR-1992) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC
CC NEURONES.
CC -1- CATALYTIC ACTIVITY: L-TYROSINE + TETRAHYDROPTERIDINE + O(2) -
CC 3,4-DIHYDROXY-L-PHENYLALANINE + DIHYDROPTERIDINE + H(2)O.
CC -1- COFACTOR: FERROUS ION.
CC -1- ENZYME REGULATION: PHOSPHORYLATION LEADS TO AN INCREASE IN THE
CC CATALYTIC ACTIVITY.
CC -1- PATHWAY: RATE-LIMITING STEP IN CATECHOLAMINE BIOSYNTHESIS.
CC -1- SIMILARITY: BELONGS TO THE BIOPTERIN-DEPENDENT AROMATIC AMINO ACID
CC HYDROXYLASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M69200; AAA40434.1; -
CC EMBL: X53503; CAA37580.1; -
CC PIR: JN0068; JN0068.
CC HSSP: P04177; ITOH.
CC MGD: MGI:98735; TH.
CC InterPro: IPR001273; -
CC Pfam: PF00351; biopterin_H; 1.
CC PRINTS: PR00372; FWHYDRXLASE.
CC PROSITE: PS00367; BIOPTERIN_HYDROXYL; 1.
CC Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;
CC Neurotransmitter biosynthesis; Phosphorylation.
CC MOD_RES 19 19 PHOSPHORYLATION (BY CAM-KINASE II)
CC (BY SIMILARITY).
CC MOD_RES 31 31 PHOSPHORYLATION (BY SIMILARITY).
CC MOD_RES 40 40 PHOSPHORYLATION (BY CAPK)
CC (BY SIMILARITY).
CC DOMAIN 51 59 POLY-ALA. SIMILARITY).
CC METAL 331 331 IRON (BY SIMILARITY).
CC METAL 336 336 IRON (BY SIMILARITY).
CC METAL 376 376 IRON (BY SIMILARITY).
CC SEQUENCE 498 AA; 55992 MW; 62790179664F6DC6 CRC64;
CC -----
Query Match 11.9%; Score 225.5; DB 1; Length 498;
Best Local Similarity 27.5%; Pred. No. 1.3e-10;
Matches 66; Conservative 42; Mismatches 101; Indels 31; Gaps 5;
QY 107 WYRLSSRFLSKSY-CPRFFLDYLEAFGLSLDFLDH-----QAVIKFELETHFSY 158
DB 223 WREYVATLKLGYATHACR---EHLEAFQLLERYCYGRSDSIPQLEDVSHLKERTGFQLR 289
QY 159 YVSGFVAPHOYLSLQDQRYFPPIASVMRTLDKDNFSLTDLIHLGLHVPMLLPSPSEFF 218
DB 290 PVAGLLSARDFLASLAFRVFOCTQYIRHASSPMHSPEDPCVHELLGHVPLMDRTTAQFS 349
QY 219 INMGRLETKVIEKQVQALPSKQRIQTLQSNLIAIVRCFWFTVESGLIENHEGRKAYGAVL 278
DB 350 QDIG-----LASLGASDEIEKLST-----VWFTVEFGCKQNGELKAYGAGL 393
QY 279 ISSPQELGHAFIDNVRVLPLELDQIIRLPNTSTPQETLFSIRHF---DELVELSKLE 334
DB 394 LSSYGLVHLSDEPERREFDPTAAVQYQDTQYQVYFVSESFSDAKDLNLYASRIQ 453
RESULT 10
TR5H_XENLA
ID TR5H_XENLA STANDARD; PRT: 481 AA.
AC Q92142;
DT 15-JUL-1998 (Rel. 36, Created)

```









ID TY3H\_DROME STANDARD; PRT; 508 AA.  
AC P18459;  
DT 01-NOV-1990 (Rel. 16, Created)  
DT 01-NOV-1990 (Rel. 16, Last sequence update)  
DT 15-JUL-1998 (Rel. 36, Last annotation update)  
DE TYROSINE 3-MONOOXYGENASE (EC 1.14.16.2) (TYROSINE 3-HYDROXYLASE) (TH)  
DE (PROTEIN PALE)  
GN TH OR PLE.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
OX NCBI\_TaxID=7227;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90166583; PubMed=2483109;  
RA Neckmeyer W.S.; Quinn W.G.;  
RA "Isolation and characterization of the gene for Drosophila tyrosine  
hydroxylase";  
RN Neuron 2:1167-1175(1989).  
CC -1- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE PHYSIOLOGY OF ADRENERGIC  
CC NEURONES.  
CC -1- CATALYTIC ACTIVITY: L-TYROSINE + TETRAHYDROPTERIDINE + O(2) =  
CC 3,4-DIHYDROXY-L-PHENYLALANINE + DIHYDROPTERIDINE + H(2)O.  
CC -1- COFACTOR: FERROUS ION.  
CC -1- PATHWAY: RATE-LIMITING STEP IN CATECHOLAMINE BIOSYNTHESIS.  
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID  
CC HYDROXYLASES FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; U14395; AAA62877.1; -  
CC EMBL; X76209; CAA53802.1; -  
CC PIR; JN0010; JN0010.  
CC HSSP; P04177; ITOH.  
CC FlyBase; FBgn0005626; ple.  
CC InterPro; IPR001273; -  
CC Pfam; PF00351; bioterin.H; 1.  
CC PRINTS; PR00372; FWHYDRXILASE.  
CC PROSITE; PS00367; BIOTERIN\_HYDROXYL; 1.  
CC Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;  
KW Neurotransmitter biosynthesis.  
FT METAL 338 338 IRON (BY SIMILARITY).  
FT METAL 343 343 IRON (BY SIMILARITY).  
FT METAL 383 383 IRON (BY SIMILARITY).  
SQ SEQUENCE 508 AA; 57861 MW; 2D194E0E8495E66F CRC64;

Query Match 11.6%; Score 219; DB 1; Length 508;  
Best Local Similarity 28.6%; Pred. No. 4.3e-10;  
Matches 74; Conservative 37; Mismatches 94; Indels 54; Gaps 8;  
QY 128 DYLEAFGLSD---FLDH-----QAVIKFELETHFSYYPVSGFVAPHOYLSDQDRYFP 179  
DB 258 EYRAAFKQLQDEIFVETRLPOLQEMSDLRKNTGSLRPAAGLLTARDLASFARIFQ 317  
QY 180 IASVARTLDKNFSLTDLHDLGHVPWLLHPFSFSEFFINMGRLFTVKVVKVQALPSKK 239  
DB 318 STQYVHVNSPYHTPEPDSITHELGHMPLVADPSFAQFSQIEG-----LASGASD 368  
QY 240 QRIOTLQSNLIAVRCVFWFVSGSLIENHEGRKAYGAVLISSPQELGHAFID----- 291  
DB 369 EIEKLST-----VWFTVFERGLCKHQIKAYGAGLSSYGGELLHAISDKCEHRAFE 421  
QY 292 --NVAVLPLE-----LDQILRLPNTPOETLFSIRHFD--ELVELT 330  
DB 422 PASTAVQYQDQEQYPIYVAESFEDAKDKFRWVSTNSRPEVRFN-PHTERVEVLDVS 480

QY 331 SKLEWMLDQGLLESIPLYN 349  
DB 481 DKLETLVHQMNTTEILHLTN 499  
RESULT 15  
TY3H\_SCHMA STANDARD; PRT; 465 AA.  
AC O17446;  
DT 15-JUL-1998 (Rel. 36, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 15-JUL-1999 (Rel. 38, Last annotation update)  
DE TYROSINE 3-MONOOXYGENASE (EC 1.14.16.2) (TYROSINE 3-HYDROXYLASE) (TH).  
GN TH.  
OS Schistosoma mansoni (Blood fluke).  
OC Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;  
OC Trematoda; Digenea; Strigeidae; Schistosomatidae; Schistosomatidae;  
OC Schistosoma.  
OX NCBI\_TaxID=6183;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=PUERTO RICAN;  
RA MEDLINE=98421787; PubMed=9751167;  
RA Hamdan F.F.; Ribeiro P.;  
RA "Cloning and characterization of a novel form of tyrosine hydroxylase  
from the human parasite, Schistosoma mansoni";  
RN J. Neurochem. 71:1369-1380(1998).  
CC -1- CATALYTIC ACTIVITY: L-TYROSINE + TETRAHYDROPTERIDINE + O(2) =  
CC 3,4-DIHYDROXY-L-PHENYLALANINE + DIHYDROPTERIDINE + H(2)O.  
CC -1- COFACTOR: FERROUS ION.  
CC -1- PATHWAY: RATE-LIMITING STEP IN CATECHOLAMINE BIOSYNTHESIS.  
CC -1- SIMILARITY: BELONGS TO THE BIOTERIN-DEPENDENT AROMATIC AMINO ACID  
CC HYDROXYLASES FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF030336; AAC62256.1; -  
CC HSSP; P04177; ITOH.  
CC InterPro; IPR001273; -  
CC Pfam; PF00351; bioterin.H; 1.  
CC PROSITE; PS00367; BIOTERIN\_HYDROXYL; 1.  
CC Catecholamine biosynthesis; Oxidoreductase; Monooxygenase; Iron;  
KW Neurotransmitter biosynthesis.  
FT METAL 294 294 IRON (BY SIMILARITY).  
FT METAL 299 299 IRON (BY SIMILARITY).  
FT METAL 339 339 IRON (BY SIMILARITY).  
SQ SEQUENCE 465 AA; 54081 MW; 0DE47B4A474665CB CRC64;  
Query Match 11.5%; Score 216.5; DB 1; Length 465;  
Best Local Similarity 24.9%; Pred. No. 6.1e-10;  
Matches 59; Conservative 49; Mismatches 88; Indels 41; Gaps 5;  
QY 128 DYLEAFGLSDFLDH-----QAVIKFELETHFSYYPVSGFVAPHOYLSDQDRYFP 179  
DB 214 EYIDGOLFGEKYCNYNSESIPOLQICEFMHRTSGFIRPVAGLVSPKFLASLAFRVFQ 273  
QY 180 IASVARTLDKNFSLPDLHDLGHVPWLLHPFSFSEFFINMGRLFTVKVVKVQALPSKK 239  
DB 274 CTQYIRHHSRPMHTPEPDCITHELGHMPLVNRQFADPSQELG-----LASGASEEI 327  
QY 240 QRIOTLQSNLIAVRCVFWFVSGSLIENHEGRKAYGAVLISSPQELGHAFIDNVRLPLE 299  
DB 328 TRLSLT-----YWFTVFERGLCKHQIKAYGAGLSSYGGELLENAP----- 367  
QY 300 LDQILRLPNTPOETLFSIRHFDVLELTSKLEWM-----LDQGLLESIPLYN 349

Mon Jul 2 08:46:44 2001

us-09-438-185-1047.rsp

Page 11

Db 368 SDLSVKEPFNINDAAVOYDDVGQKIYFVTESIESMKRELNYINTSGKSTIPIYD 424

Search completed: June 29, 2001, 13:56:09  
Job time: 238 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2001, 13:52:51 ; Search time 13.54 Seconds  
(without alignments)  
538.579 Million cell updates/sec

Title: US-09-438-185-1047

Perfect score: 1889

Sequence: 1 VHYCERTLPKYLKALKL.....ESIPLYNQEKYLSGFVLQ 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 193259 seqs, 2014635 residues

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A-COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B-COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A-COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B-COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/ECTUS-COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	% Match	Query Length	DB ID	Description
1	223	11.8	498	1	US-08-009-075-5
2	220	11.6	497	1	US-08-009-075-4
3	98.5	5.2	1024	2	US-09-091-117-5
4	95	5.0	1226	2	US-08-540-804-12
5	95	5.0	1226	2	US-08-218-265-12
6	95	5.0	1226	3	US-08-321-872-12
7	95	5.0	1226	3	US-08-590-399-12
8	91.5	4.8	872	1	US-08-766-014-2
9	91	4.8	535	3	US-08-369-822C-24
10	91	4.8	535	3	US-08-582-776C-39
11	91	4.8	535	4	US-08-434-831B-36
12	91	4.8	2183	2	US-08-348-891A-7
13	91	4.8	2183	2	US-08-905-817-7
14	90.5	4.8	615	1	US-08-484-105-6
15	90.5	4.8	615	1	US-08-484-106-6
16	89	4.7	392	3	US-08-911-853-35
17	89	4.7	392	4	US-09-479-409-35
18	87.5	4.6	309	3	US-08-996-338-22
19	87.5	4.6	521	3	US-08-996-338-20
20	87.5	4.6	541	3	US-08-604-333-2
21	87.5	4.6	541	3	US-09-110-618-2
22	86.5	4.6	990	2	US-08-392-625-20
23	86.5	4.6	990	2	US-08-466-963A-20
24	86.5	4.6	990	2	US-08-466-963B-20
25	83.5	4.4	1365	6	US-08-845-193B-15
26	83	4.4	491	2	5194600-4
27	82	4.3	861	3	US-08-942-819-2
					US-09-022-875-2

28	81.5	4.3	608	1	US-08-766-014-3	Sequence 3, Appli
29	81.5	4.3	1708	1	US-08-493-092-2	Sequence 2, Appli
30	81.5	4.3	1708	1	US-08-508-836A-2	Sequence 2, Appli
31	81.5	4.3	3056	1	US-08-508-836A-8	Sequence 8, Appli
32	81.5	4.3	3056	2	US-08-629-001A-3	Sequence 3, Appli
33	81.5	4.3	3056	2	US-08-874-266-2	Sequence 2, Appli
34	81.5	4.3	3056	2	US-08-642-274D-3	Sequence 3, Appli
35	81.5	4.3	3056	4	US-08-952-127-3	Sequence 3, Appli
36	81	4.3	323	2	US-08-540-804-18	Sequence 18, Appli
37	81	4.3	323	2	US-08-218-265-18	Sequence 18, Appli
38	81	4.3	323	3	US-08-521-872-18	Sequence 18, Appli
39	81	4.3	323	4	US-08-590-399-18	Sequence 18, Appli
40	81	4.3	474	4	US-08-961-083-60	Sequence 60, Appli
41	81	4.3	1118	2	US-08-724-354D-2	Sequence 2, Appli
42	81	4.3	1118	3	US-09-270-984A-2	Sequence 2, Appli
43	80.5	4.3	377	1	US-07-863-169A-1	Sequence 1, Appli
44	80.5	4.3	377	2	US-08-429-964-1	Sequence 1, Appli
45	80.5	4.3	377	3	US-07-935-087-1	Sequence 1, Appli

## ALIGNMENTS

RESULT 1  
US-08-009-075-5  
; Sequence 5, Application US/08009075  
; Patent No. 5300436  
; GENERAL INFORMATION:  
; APPLICANT: GOLDSTEIN, Menek  
; APPLICANT: WU, Jing  
; APPLICANT: FRIEDHOFF, Arnold J.  
; TITLE OF INVENTION: GENETICALLY MODIFIED TYROSINE  
; HYDROXYLASE AND USES THEREOF  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/009,075  
; FILING DATE: 19930126  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: TOWNSEND, GUY K.  
; REGISTRATION NUMBER: 34,033  
; REFERENCE/DOCKET NUMBER: GOLDSTEIN=IA  
; TELEPHONE: 202-628-5197  
; TELEFAX: 202-737-3528  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 498 amino acids  
; TYPE: AMINO ACID  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-009-075-5

Query Match 11.8%; Score 223; DB 1; Length 498;  
Best Local Similarity 26.7%; Pred. No. 9.9e-16;  
Matches 65; Conservative 39; Mismatches 103; Indels 36; Gaps 5;

Qy 112 SSRFLSKWS-----YCPRFLLDYLEAFGLSLDFLDH-----QAVIKFFLETHF 155

Db 227 ABEIATWKEVTVTLKGLYATHACREHLEGQLLERYCGYREDSPQLEDVSRFLKERTGF 286  
QY 156 SYYPVSGFVAPHOVLSLQDRYFPPIASVMRTLDKNFSLTPDLIHDLLGHVPLLLHPFSFS 215  
Db 287 QLRPVAGLLSARDFLASLAFRVFQCTQYIRHASSPMHSPEDPCCHELLGHVPMPLADRTFA 346  
QY 216 EFFINMGRLEFTKVIKQALPSKORIQTLQSNLAIIVRCFWFTVESGLIENHGRKAYG 275  
Db 347 QFSODIG-----LASIGASDEIEKLSLTS-----VWFTVEFGCLCKQNGELKAYG 390  
QY 276 AVLISSPOELGHAFIDNVRLPLEDOIIRLPNTSTPOETLFSIRHF-----DELVELTS 331  
Db 391 AGLSSYGELLHSLSEPEVRAFDPTAAVQPDQTYQFVYVSESFNDAKDLRNYAS 450  
QY 332 KLE 334  
451 RIQ 453

RESULT 2  
US-08-009-075-4  
; Sequence 4, Application US/08009075  
; Patent No. 5300436  
; GENERAL INFORMATION:  
; APPLICANT: GOLDSTEIN, Menek  
; APPLICANT: WU, Jinq  
; APPLICANT: FRIEDHOFF, Arnold J.  
; TITLE OF INVENTION: GENETICALLY MODIFIED TYROSINE  
; TITLE OF INVENTION: HYDROXYLASE AND USES THEREOF  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BROWDY AND NEIMARK  
; STREET: 419 Seventh Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: U.S.A.  
; ZIP: 20004  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/009,075  
; FILING DATE: 19930126  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: TOWNSEND, GUY K.  
; REGISTRATION NUMBER: 34,033  
; REFERENCE/DOCKET NUMBER: GOLDSTEIN-1A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-628-5137  
; TELEFAX: 202-737-3528  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 497 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-009-075-4

Query Match 11.6%; Score 220; DB 1; Length 497;  
Best Local Similarity 26.7%; Pred. No. 2.1e-15;  
Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;  
QY 112 SRRFSLWKS-----YCPRFEDLYEAFGLLSDFLDH-----QAVLKFFELETHF 155  
Db 226 ABEIATWKEVTVTLKGLYATHACREHLEGQLLERYCGYREDSPQLEDVSRFLKERTGF 285  
QY 156 SYYPVSGFVAPHOVLSLQDRYFPPIASVMRTLDKNFSLTPDLIHDLLGHVPLLLHPFSFS 215

Db 286 QLRPVAGLLSARDFLASLAFRVFQCTQYIRHASSPMHSPEDPCCHELLGHVPMPLADRTFA 345  
QY 216 EFFINMGRLEFTKVIKQALPSKORIQTLQSNLAIIVRCFWFTVESGLIENHGRKAYG 275  
Db 346 QFSODIG-----LASIGASDEIEKLSLTS-----WFTVEFGCLCKQNGEVKAYG 389  
QY 276 AVLISSPOELGHAFIDNVRLPLEDOIIRLPNTSTPOETLFSIRHF-----DELVELTS 331  
Db 391 AGLSSYGELLHSLSEPEVRAFDPTAAVQPDQTYQFVYVSESFNDAKDLRNYAS 449  
QY 332 KLE 334  
451 RIQ 453

RESULT 3  
US-09-091-117-5  
; Sequence 5, Application US/09091117  
; Patent No. 6171589  
; GENERAL INFORMATION:  
; APPLICANT: The University of Melbourne  
; TITLE OF INVENTION: Mycoplasma Recombinant Polypeptides and  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GREENLEE, WINNER and SULLIVAN P.C.  
; STREET: 5370 Manhattan Circle, Suite 201  
; CITY: Boulder  
; STATE: Colorado  
; COUNTRY: United States of America  
; ZIP: 80303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/091,117  
; FILING DATE: 12 JUNE 1998  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/AU96/00803  
; FILING DATE: 13-DEC-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: AU PN7127  
; FILING DATE: 13-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WINNER, Ellen P.  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: +1 303 499 8080  
; TELEFAX: +1 303 499 8089  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1024 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Mycoplasma genitalium  
US-09-091-117-5

Query Match 5.2%; Score 98.5; DB 4; Length 1024;  
Best Local Similarity 19.9%; Pred. No. 0.091;  
Matches 68; Conservative 41; Mismatches 134; Indels 99; Gaps 11;  
QY 29 QNSQSLQAYSTPYSYRIILOKENKEK---QALAHKICISILEFFKNLLFVHLLSLKN 85  
Db 564 QOTDSLKNLFSV---IGDILSETNVNKTTHAVKNELLSLVETASTLKIKHL----- 613  
QY 86 QRECSCTDMAVSTPFPFNRLNWLRLSSRFSLWKSQCPRFEDLYEAFGLLSDFLDHQA 145  
Db 614 -----NVOYKVLVDKFKELNSFIK-----ELNFFPDTKDITPT 647



QY 146 IK--PFELETH-----ESYVPVSGVAPHOYLSLQDRYFPFIASVMTLDKDNF 192  
DB 648 IKVLESENYTKKKYENEGPGYHWAKFVPGTFNSAENFTFSAL-----DKT 698  
QY 193 SLTPDLIHLHGHVPLWHPSEFFINMGRLETF-----KVIEKVOALPSKKQRIQT 244  
DB 699 KSTRDLFADML--FGKLSVANDSDSFIKNGSTLKYHGDNLNLLPNYHSLITKNVGYOI 757  
QY 245 LQSNLIAVRCFWFTVESGLIENHEGRKAYGAVLISPOELG-----286  
DB 758 VVNFHIDARLLTAELONTVFSNPK-----PVIKSPVELSKSLFEVWKTFIENSNOI 810  
QY 287 ---HAFIDNVRVPLELDQIIRLPFNFTSTPQETLFSIRHFD 324  
DB 811 LKKEYTFKDNLFKFFPKAGDSRLEFDSKPDQVRVPPAFVD 852

## RESULT 4

US-08-540-804-12  
; Sequence 12, Application US/08540804  
; Patent No. 5919666  
; GENERAL INFORMATION:  
; APPLICANT: Young, Richard A.  
; APPLICANT: Koleske, Anthony J.  
; APPLICANT: Thompson, Craig M.  
; APPLICANT: Chao, David M.  
; TITLE OF INVENTION: No. 5919666el Factors Which Modify Gene  
; TITLE OF INVENTION: Transcription and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/540,804  
FILING DATE: 11-OCT-1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/521,872  
FILING DATE: 21-AUG-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/218,265  
FILING DATE: 25-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: WHI94-03A2  
TELEPHONE: 617-861-9540  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
TYPE: amino acid  
LENGTH: 1226 amino acids  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-540-804-12

Query Match 5.08; Score 95; DB 2; Length 1226;  
Best Local Similarity 18.8%; Pred. No. 0.29;  
Matches 98; Conservative 65; Mismatches 155; Indels 204; Caps 25;  
QY 12 YILKIALKRLQSLFFQNSQLORAYSTPYRYII-----LOKENKQOALAHKCI 65

DB -13 YILE--KLIPDMTHYNDSQL-RTWKROIISYFLKLLGNCYSRLINKE--IPHHWLV 65  
QY 66 SILEFFKNLFF-----VHLLSLSKNOREGSCSTDMAYST-----PPF-----102  
DB 66 EFINKMENEFPLSLHILMIFWMDICOIDTINAPVAATITSSQKEPFFLVTKITDMLHK 125  
QY 103 -----NRNLWYRLSSRFSL-----WKS 120  
DB 126 YIIVSSKSMINDENYIINDIKKNNKIKLNLKILSLILKIFQEOSLEVFFPPTSNNWEI 185  
QY 121 YCPREF-----LDYLEAFGLISDFLDHQAVIKFFTELETHFSYYPVSGFVAP 166  
DB 186 YKPLLEIVSNADTNQNSDMKKKLELISYRNESLKNNSIR-----NVIMSASNAN 236  
QY 167 HQVLSLQDRYFPFIASV-----RTLDKDN-----FSLTPDLIHLHGHVP 207  
DB 237 DFQLTIVTCKQPKLSCIQLNCIDTQTKLLD-DNPTFDWPTYVDQNLTMHKIQLIL 295  
QY 208 WLLHPS--PSEFFINMGRLETFKVEKVOALPSKKORIOTLOSNI--IAIVRCFWFTVESG 263  
DB 296 WSHPSRQFDHYESN-----QLVAKLLLL-----RINSTDEDLHEFQIEDAINSLVFO- 343  
QY 264 LIENHEGRK-----AYGAV-----LISSP-----OELGHAFID-- 291  
DB 344 LAKNSAQKRVSYWMPSLYRLNLITYGIKVPYIRKLLISSGLLYLQDSNDRKFVHVQ 403  
QY 292 ---NVRVPLELDQIIRLPFNFTSTPQETLFSIRHDELVELTSL-----333  
DB 404 LLINKISPLMKSOYNMVLNRNMEYDKVFEINFQDOLVEITQIKMKRILSDITNLQLS 463  
QY 334 -----EWMLD--OGLLESIPLYNQEKYLSGFVLC 361  
DB 464 KTPLSKIMVAEYLSHLCSGILSV--NRTVLLKTFKIFC 502

## RESULT 5

US-08-218-265-12  
; Sequence 12, Application US/08218265  
; Patent No. 5922585  
; GENERAL INFORMATION:  
; APPLICANT: Young, Richard A.  
; APPLICANT: Koleske, Anthony J.  
; APPLICANT: Thompson, Craig M.  
; TITLE OF INVENTION: No. 5922585el Factors Which Modify Gene  
; TITLE OF INVENTION: Transcription and Methods of Use Thereof  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02173  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/218,265  
FILING DATE: 25-MAR-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Granahan, Patricia  
REGISTRATION NUMBER: 32,227  
REFERENCE/DOCKET NUMBER: WHI94-03  
TELEPHONE: 617-861-9540  
TELEFAX: 617-861-9540  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1226 amino acids

; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-218-265-12

Query Match 5.0%; Score 95; DB 2; Length 1226;  
Best Local Similarity 18.8%; Pred. No. 0.29;  
Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;

QY 12 YILKIALKRLQSLFQNSQSLQRAYSTPYSYRII-----LOKENKEQALARKCI 65  
DB 13 YILE---KLIFDMNHYNDSQOL-RTWKQISYFLKLGNCYSRLINKE---IFHWLV 65  
QY 66 SILEFFKNLLF---VHLLSLSKNOREGCGSTDMAVST-----PFF----- 102  
DB 66 EFINKMENFEFLPLSLHILMFNDICQIDTINAPVAATITSSQKEPFLVTKITDMLLHK 125  
QY 103 -----NRNLWYLLSSRFSL-----WKS 120  
DB 126 YIVSSSKSMINDENYIINDIKNNKIKILNLKILSSILKIFQOSLEVFIFPTSNEI 185  
QY 121 YCPREF-----LDYLEAFGLSDPLDHOAVIKFEFELETHFSYYPVSGFVAP 166  
DB 186 YKPLLEIVSNADTNQNSDKMKKLELISYRNESLKNSSIR-----NVIMSASNAN 236  
QY 167 HOYLSLLODRFFPIASV-----RTLDKN-----FSLTPDLIHLDLGHVP 207  
DB 237 DPQLTIVTCKQFPKLSQIQLNCIDTOFTKLLD-DNPTFDMPTVYDQNLPMHKKIQLIL 295  
QY 208 WLLHPS---FSEFFINMGRLETKVIEKQALPSKKQRIOTLOSNI--IAIVRCFWFTVESG 263  
DB 296 WSIHPSRQFDHYESN-----QLVAKLLL-----RINSTDEDLHEFOIEDAIVSLVFQ- 343  
QY 264 LIENHEGRK-----AYGAV-----LISSP-----OELGHAFID-- 291  
DB 344 LAKNFSACKRVVSYMPSLYRLNLITLITYGIKVPYIRKLSSGLLYLQDSNDFVHVQ 403  
QY 292 ---NVRVLPLELDQIIRLPNTSTPQETLFSIRHFDLVELTSKL----- 333  
DB 404 LLINKISPLMKSOYNKVLNRVMEYDKVFEYFNFDQVLEITEQIKMRILSDITNLQLS 463  
QY 334 -----EWMLD---OGLESPLYNQEKYLSGFVLC 361  
DB 464 KTPLSIKIMVAEWYLSHLCSGILSSV---NRTVLLKIFKIFC 502

## RESULT 6

US-08-521-872-12  
; Sequence 12, Application US/08521872  
; Patent No. 6015682  
; GENERAL INFORMATION:  
; APPLICANT: Young, Richard A.  
; APPLICANT: Koleske, Anthony J.  
; APPLICANT: Thompson, Craig M.  
; APPLICANT: Chao, David M.  
; TITLE OF INVENTION: No. 6015682el Factors Which Modify Gene  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Willita Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/521,872

; FILING DATE: 31-AUG-1995  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/218,265  
; FILING DATE: 25-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: WHI94-03A  
; TELEPHONE: 617-861-9540  
; TELEFAX: 617-861-6240  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1226 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-521-872-12

Query Match 5.0%; Score 95; DB 3; Length 1226;  
Best Local Similarity 18.8%; Pred. No. 0.29;  
Matches 98; Conservative 65; Mismatches 155; Indels 204; Gaps 25;

QY 12 YILKIALKRLQSLFQNSQSLQRAYSTPYSYRII-----LOKENKEQALARKCI 65  
DB 13 YILE---KLIFDMNHYNDSQOL-RTWKQISYFLKLGNCYSRLINKE---IFHWLV 65  
QY 66 SILEFFKNLLF---VHLLSLSKNOREGCGSTDMAVST-----PFF----- 102  
DB 66 EFINKMENFEFLPLSLHILMFNDICQIDTINAPVAATITSSQKEPFLVTKITDMLLHK 125  
QY 103 -----NRNLWYLLSSRFSL-----WKS 120  
DB 126 YIVSSSKSMINDENYIINDIKNNKIKILNLKILSSILKIFQOSLEVFIFPTSNEI 185  
QY 121 YCPREF-----LDYLEAFGLSDPLDHOAVIKFEFELETHFSYYPVSGFVAP 166  
DB 186 YKPLLEIVSNADTNQNSDKMKKLELISYRNESLKNSSIR-----NVIMSASNAN 236  
QY 167 HOYLSLLODRFFPIASV-----RTLDKN-----FSLTPDLIHLDLGHVP 207  
DB 237 DPQLTIVTCKQFPKLSQIQLNCIDTOFTKLLD-DNPTFDMPTVYDQNLPMHKKIQLIL 295  
QY 208 WLLHPS---FSEFFINMGRLETKVIEKQALPSKKQRIOTLOSNI--IAIVRCFWFTVESG 263  
DB 296 WSIHPSRQFDHYESN-----QLVAKLLL-----RINSTDEDLHEFOIEDAIVSLVFQ- 343  
QY 264 LIENHEGRK-----AYGAV-----LISSP-----OELGHAFID-- 291  
DB 344 LAKNFSACKRVVSYMPSLYRLNLITLITYGIKVPYIRKLSSGLLYLQDSNDFVHVQ 403  
QY 292 ---NVRVLPLELDQIIRLPNTSTPQETLFSIRHFDLVELTSKL----- 333  
DB 404 LLINKISPLMKSOYNKVLNRVMEYDKVFEYFNFDQVLEITEQIKMRILSDITNLQLS 463  
QY 334 -----EWMLD---OGLESPLYNQEKYLSGFVLC 361  
DB 464 KTPLSIKIMVAEWYLSHLCSGILSSV---NRTVLLKIFKIFC 502

## RESULT 7

US-08-590-399-12  
; Sequence 12, Application US/08590399  
; Patent No. 6214588  
; GENERAL INFORMATION:  
; APPLICANT: Young, Richard A.  
; APPLICANT: Koleske, Anthony J.  
; APPLICANT: Thompson, Craig M.  
; APPLICANT: Chao, David M.  
; TITLE OF INVENTION: No. 6214588el Factors Which Modify Gene  
; TITLE OF INVENTION: Transcription and Methods of Use Therefor



Db 310 TLQD-----NNIAFYPLIYEIGEIKKIAFGKDTVYIDVFQTE-----DL 351  
QY 184 MRTLKDNFSLTPDLIHLA-----GHVPWLLHPSFSEFFINWGLFT 226  
Db 352 KEIFKEDEFETHEIKDFLVRISYKIGIECKSKYIDTAVMAYLLMPSESNT--DLDRVLK 409  
QY 227 KVIKQALPS-----KKORIOTLOSNIUA-----IVRCFW----- 257  
Db 410 KVL-KVD-VPSYEGIFGKGKRDKKIEIDENILADYICSRVYLFDLKEKIMNFIERMDM 467  
QY 258 ----FTVESGLIENHEGRKAYGAVLJSSP-QELGHAFIDNVRVLPLELDQIIR---LPFN 309  
Db 468 KLLLEIEMPIVEVLKMEYSGFTLDEKVLKLSQKIDDRIGEI---LDKIYKEAGYQFN 524  
QY 310 TSTPOE-TLFSIRHFDLVELTSLKLEWMLDQGLLESIPLYN-----QEKY 353  
Db 525 VNSPKQLSEFLKLNLPVKKTKGTGYSTDSEVLQVLPYNDIVSDIIEYRQLTKLKSTY 584  
QY 354 LSGF 357  
Db 585 IDGF 588

RESULT 9  
US-08-369-822C-24  
; Sequence 24, Application US/08369822C  
; Patent No. 6015860  
; GENERAL INFORMATION:  
; APPLICANT: Lipkin, W. I.  
; APPLICANT: Briese, Thomas  
; APPLICANT: Kliche, Stefanie  
; APPLICANT: Schneider, Patrick A.  
; APPLICANT: Stitz, Lothar  
; APPLICANT: Schneemann, Anette  
; TITLE OF INVENTION: Borna Disease Viral Sequences,  
; TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous  
; TITLE OF INVENTION: System Diseases  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fulbright & Jaworski, L.L.P.  
; STREET: 865 South Figueroa Street, 29th Floor  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90017-2571  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0  
; SOFTWARE: ASCII (DOS) TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/369,822C  
; FILING DATE: 06-JAN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Churchill, Margaret A. (Ph.D.)  
; REGISTRATION NUMBER: 39,944  
; REFERENCE/DOCKET NUMBER: 1279-194XX  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 213/892-9200  
; TELEFAX: 213/680-4518  
; INFORMATION FOR SEQ ID NO: 24:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 535 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-369-822C-24

Query Match 4.8%; Score 91; DB 3; Length 535;  
Best Local Similarity 20.6%; Pred No. 0.22; Indels 136; Gaps 18;  
Matches 78; Conservative 46; Mismatches 119;  
QY 57 QALARKHCISILEFFKNLLFVHLLSLSKNQREGGCTDMVVSTPFFNNRWYLLSSRFS 116  
Db 1 EGLTHQOCVDNWKSPAGVAF-----GCFMPLSLDSD--LTMYLKRALAALQR 46  
QY 117 LKWSYCPREFLDYLEAGLLSDFLDHQAVIKFELETHFSYYPVSGFVAPHQYLSLQDR 176  
Db 47 EWSVYKPEFLRYDPDPKGTGS-----RRLVDVFLNDSFDPYDVMYVVSAGY---LHDP 98  
QY 177 YFTASVMRTLDKDNESLTPDLIHLHGVFWLLHPSFSEFFINWGLFTKVIKQALP 236  
Db 99 EFNL-----SYSLOEKEIKE-----TGRLEAKMTYKMR-- 127  
QY 237 SKQRTQTLOSNIUA-----IVRCFWFTVESG---LIENHEGRKAY 274  
Db 128 ----COVIAENLISNGIKGYKDNMGMAKDEQDLTKALHTLAVSGVPKDKESHG--- 178  
QY 275 GAVLISSPOELGHAFIDNVRVLP--LELDQIIRLPENTSTPOE----- 315  
Db 179 GVLKTYRSRPHVTSTRNVRRAKGFIFGFPQVIRQDDTDHPENMEAYETVSAFITDLKK 238  
QY 316 -----TLFSIRHFDLVELTSLKLEWMLDQGLL-----ESIPLY-- 348  
Db 239 YCLNRYETISLFAQR-LNEIYGLPSFFQWLHKLRLTSVLVSDPHCPDLDLHIEPLYK 297  
QY 349 -NOE---KY-LSGFVFLCQ 362  
Db 298 PNDQIFIKYPMGIGEGYCQ 316  
RESULT 10  
US-08-582-776C-39  
; Sequence 39, Application US/08582776C  
; Patent No. 6077510  
; GENERAL INFORMATION:  
; APPLICANT: Lipkin, W. I.  
; APPLICANT: Briese, Thomas  
; APPLICANT: Kliche, Stefanie  
; APPLICANT: Schneider, Patrick A.  
; APPLICANT: Stitz, Lothar  
; APPLICANT: Schneemann, Anette  
; TITLE OF INVENTION: Borna Disease Viral Sequences,  
; TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous  
; TITLE OF INVENTION: System Diseases  
; NUMBER OF SEQUENCES: 61  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fulbright & Jaworski, L.L.P.  
; STREET: 865 South Figueroa Street, 29th Floor  
; CITY: Los Angeles  
; STATE: California  
; COUNTRY: USA  
; ZIP: 90017-2576  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: WINDOWS NT  
; SOFTWARE: ASCII DOS TEXT  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/582,776C  
; FILING DATE: 04-JAN-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/369,822  
; FILING DATE: 06-JAN-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/434,831  
; FILING DATE: 04-MAY-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Churchill, Margaret A.  
; REGISTRATION NUMBER: 39,944

REFERENCE/DOCKET NUMBER: 1279-194C2  
TELEPHONE: 213/892-9200  
TELEFAX: 213/680-4518  
INFORMATION FOR SEQ ID NO: 39:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 535 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-582-776C-39

Query Match 4.8%; Score 91; DB 3; Length 535;  
Best Local Similarity 20.6%; Pred. No. 0.22;  
Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;

QY 57 QALARKHCISILEFFKLLFVHLLSLKNOEGCSTDMAVSTPFNNLWYRLSSRFS 116  
DB 1 EGLTHQCCVDNKKSFAGVKF-----GCFMPLSLDS--LTMYLKDKAALALOR 46

QY 117 LKMSYCPREFLDYLEAFGLSLDFLDHQAVIKFLELTHFSYYPVSGFVAPHQYLSLQDR 176  
DB 47 EWSVYPKREFLRDPKGTGS-----RLVDVFLNDSDFDYDVIWVVGAY---LHDP 98

QY 177 YFPIASVMTLKDKNFSITPDLIHDLGHVPMLLHPSFEFFINMGRLETKVIEKVQALP 236  
DB 99 EFNL-----SYSLOEKEIKE-----IVRCFWFTVESG---LIENHEGRKAY 127

QY 237 SKKQRIQTLOSNLIA-----CQVIAENLSNGIGYKFGKONGMAKDEQDLTKALHTLAVSGVPKDKESHG 274  
DB 128 -----CQVIAENLSNGIGYKFGKONGMAKDEQDLTKALHTLAVSGVPKDKESHG 178

QY 275 GAVLISSPOELGHAFIDNVRVLP--LELDQIIRLPNTSTPQE----- 315  
DB 179 GPKVLYSRSPVHTSTRNVRRAAKGFIGFPQVIRQDQDTHPENMEAYETVSAFTTDLKK 238

QY 316 -----TLFSIRHDFELVELTSKLEW---LDOGLL-----ESIPLY-- 348  
DB 239 YCLNWRVETISLFAQR-LNEIYGLPSFFQWLHKLRLTSLVLYSDPHCPDLDLDAHIPLYK 297

QY 349 -NOE---KY-LSGFEVLQ 362  
DB 298 PNDQIFIKYPMGGIEGYCQ 316

RESULT 11  
US-08-434-831B-36  
Sequence 36, Application US/08434831B  
Patent No. 6113905  
GENERAL INFORMATION:  
APPLICANT: Lipkin, W. I.  
APPLICANT: Briesse, Thomas  
APPLICANT: Kliche, Stefanie  
APPLICANT: Schneider, Patrick A.  
APPLICANT: Stitz, Lothar  
APPLICANT: Schneemann, Anette  
TITLE OF INVENTION: Borna Disease Viral Sequences  
TITLE OF INVENTION: Diagnostics and Therapeutics for Central Nervous  
TITLE OF INVENTION: System Diseases  
NUMBER OF SEQUENCES: 39  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fulbright & Jaworski, L.L.P.  
STREET: 865 South Figueroa Street, 29th Floor  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA  
ZIP: 90017-2571  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: WINDOWS NT-WORDPERFECT 8.0  
SOFTWARE: ASCII (DOS) TEXT  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/434,831B  
FILING DATE: 04-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/369,822  
FILING DATE: 06-JAN-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Churchill, Margaret A. (Ph.D.)  
REGISTRATION NUMBER: 39,944  
REFERENCE/DOCKET NUMBER: 1279-194C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 213/892-9200  
TELEFAX: 213/680-4518  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 535 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-434-831B-36

Query Match 4.8%; Score 91; DB 4; Length 535;  
Best Local Similarity 20.6%; Pred. No. 0.22;  
Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;

QY 57 QALARKHCISILEFFKLLFVHLLSLKNOEGCSTDMAVSTPFNNLWYRLSSRFS 116  
DB 1 EGLTHQCCVDNKKSFAGVKF-----GCFMPLSLDS--LTMYLKDKAALALOR 46

QY 117 LKMSYCPREFLDYLEAFGLSLDFLDHQAVIKFLELTHFSYYPVSGFVAPHQYLSLQDR 176  
DB 47 EWSVYPKREFLRDPKGTGS-----RLVDVFLNDSDFDYDVIWVVGAY---LHDP 98

QY 177 YFPIASVMTLKDKNFSITPDLIHDLGHVPMLLHPSFEFFINMGRLETKVIEKVQALP 236  
DB 99 EFNL-----SYSLOEKEIKE-----IVRCFWFTVESG---LIENHEGRKAY 127

QY 237 SKKQRIQTLOSNLIA-----CQVIAENLSNGIGYKFGKONGMAKDEQDLTKALHTLAVSGVPKDKESHG 274  
DB 128 -----CQVIAENLSNGIGYKFGKONGMAKDEQDLTKALHTLAVSGVPKDKESHG 178

QY 275 GAVLISSPOELGHAFIDNVRVLP--LELDQIIRLPNTSTPQE----- 315  
DB 179 GPKVLYSRSPVHTSTRNVRRAAKGFIGFPQVIRQDQDTHPENMEAYETVSAFTTDLKK 238

QY 316 -----TLFSIRHDFELVELTSKLEW---LDOGLL-----ESIPLY-- 348  
DB 239 YCLNWRVETISLFAQR-LNEIYGLPSFFQWLHKLRLTSLVLYSDPHCPDLDLDAHIPLYK 297

QY 349 -NOE---KY-LSGFEVLQ 362  
DB 298 PNDQIFIKYPMGGIEGYCQ 316

RESULT 12  
US-08-348-891A-7  
Sequence 7, Application US/08348891A  
Patent No. 5654136  
GENERAL INFORMATION:  
APPLICANT: SASAKI, Keiko  
APPLICANT: MORI, Takayuki  
APPLICANT: MAKINO, Satoshi  
TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE,  
TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR  
TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:

ADDRESSEE: YOUNG & THOMPSON  
STREET: 745 South 23rd Street  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION NUMBER: US/08/348,891A  
FILING DATE: 25-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
FILING DATE: 10-MAR-1992  
APPLICATION NUMBER: US 07/848,400  
PRIOR APPLICATION DATA:  
FILING DATE: 10-MAR-1992  
APPLICATION NUMBER: JP 3-293625  
FILING DATE: 14-OCT-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: PATCH, Andrew J.  
REGISTRATION NUMBER: 32,925  
REFERENCE/DOCKET NUMBER: KP-7501  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-521-2297  
TELEFAX: 703-685-0573  
TELEX: 248425 EMBON  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2183 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-348-891A-7

Query Match 4.8%; Score 91; DB 1; Length 2183;  
Best Local Similarity 20.6%; Pred. No. 1.9;  
Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;

QY 57 QALARKKCSILEFFKNLLFVHLLSKNQREGCSTDMVSTPTFFNRLMYRLSSRFS 116  
DB 429 EGLTHQOCVDNWKSFAGVKF-----GCFMPLSLDSD--LTMYLKDKALAAALQR 474  
QY 117 LKWSYCPREFLDYLEAFGLSLDFLDHQAVIKFFETHFYSYVSGFVAPHQYLSLQDR 176  
DB 475 EWDVYKPEFLRYDPPKGTGS-----RRLVDVFLNDSFDPDYIMVYVSGAY---LHDP 526  
QY 177 YFPIASVMTLKDKNFSLTDLIHLGLHVPWLLHPSFSEFFINMGRLETKVIEKVQALP 236  
DB 527 ENL-----SYSLKEKE-----TGRLEAKMTYKRA-- 555  
QY 237 SKKQRIQTLOSNI-----IVRCFWFTVESG-----LIENHEGRKAY 274  
DB 556 -----CQVIAENLISNGIKYFKDNGMAKDEHDLTKALHTLAVSGVPDKLKHSHRG-- 606  
QY 275 GAVLISSPOELGHAFIDNVRVLP--LELDQIIRLPNTSTPQE----- 315  
DB 607 GVLKTSRSVPVHTSTRNRAAKGFTGFQVIRQDQDDHPENMEAYETVSFAITTDLKK 666  
QY 316 -----TLFSIRHFDLVELTSKLEWM--LDQGL-----ESIPLY-- 348  
DB 667 YCLNWRVETISLFAQR--LNEIYGLPSFFQWLKRLTSLVYSDPHCPDOLDAIPLYKV 725  
QY 349 -NQE---KY--LSGFEVLQ 362  
DB 726 PNDQIFIKYPMGGIEGYC 744

RESULT 13  
US-08-905-817-7

Sequence 7, Application US/08905817  
Patent No. 5824777  
GENERAL INFORMATION:  
APPLICANT: SASAKI, Keiko  
APPLICANT: MORI, Takayuki  
APPLICANT: MAKINO, Satoshi  
TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE.  
TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR  
TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: YOUNG & THOMPSON  
STREET: 745 South 23rd Street  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/905,817  
FILING DATE: 04-AUG-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
FILING DATE: 25-NOV-1994  
APPLICATION NUMBER: US 08/348,891  
PRIOR APPLICATION DATA:  
FILING DATE: 10-MAR-1992  
APPLICATION NUMBER: US 07/848,400  
PRIOR APPLICATION DATA:  
FILING DATE: 14-OCT-1991  
APPLICATION NUMBER: JP 3-293625  
ATTORNEY/AGENT INFORMATION:  
NAME: PATCH, Andrew J.  
REGISTRATION NUMBER: 32,925  
REFERENCE/DOCKET NUMBER: KP-7501A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-521-2297  
TELEFAX: 703-685-0573  
TELEX: 248425 EMBON  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2183 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-905-817-7

Query Match 4.8%; Score 91; DB 2; Length 2183;  
Best Local Similarity 20.6%; Pred. No. 1.9; Indels 136; Gaps 18;  
Matches 78; Conservative 46; Mismatches 119; Indels 136; Gaps 18;  
QY 57 QALARKKCSILEFFKNLLFVHLLSKNQREGCSTDMVSTPTFFNRLMYRLSSRFS 116  
DB 429 EGLTHQOCVDNWKSFAGVKF-----GCFMPLSLDSD--LTMYLKDKALAAALQR 474  
QY 117 LKWSYCPREFLDYLEAFGLSLDFLDHQAVIKFFETHFYSYVSGFVAPHQYLSLQDR 176  
DB 475 EWDVYKPEFLRYDPPKGTGS-----RRLVDVFLNDSFDPDYIMVYVSGAY---LHDP 526  
QY 177 YFPIASVMTLKDKNFSLTDLIHLGLHVPWLLHPSFSEFFINMGRLETKVIEKVQALP 236  
DB 527 ENL-----SYSLKEKE-----TGRLEAKMTYKRA-- 555  
QY 237 SKKQRIQTLOSNI-----IVRCFWFTVESG-----LIENHEGRKAY 274  
DB 556 -----CQVIAENLISNGIKYFKDNGMAKDEHDLTKALHTLAVSGVPDKLKHSHRG-- 606  
QY 275 GAVLISSPOELGHAFIDNVRVLP--LELDQIIRLPNTSTPQE----- 315

Db 607 GPKYTSRSPVHTSTRNVRRAAKGFIGPQVIRQDQDTPDRPENMEAYETVSFAITTDLKK 666  
QY 316 -----TLFSIRHDELVELTSKLEWM---LDOGLL-----ESTPLY-- 348  
Db 667 YCLNWRYETISLSFAQR-LNEYIGLPSFPQWLHKRLKLETSVLVYSDPHCPDLDLHPIKYV 725  
QY 349 -NOE---KY-LSGFEVLQ 362  
Db 726 PNDQIFIKYPMGGIEGYCQ 744

RESULT 14  
US-08-484-105-6

; Sequence 6, Application US/08484105  
; Patent No. 5589341

; GENERAL INFORMATION:

; APPLICANT: STILLMAN, Bruce

; APPLICANT: BELL, Stephen P

; APPLICANT: KOBAYASHI, Ryuj1

; APPLICANT: RINE, Jasper

; APPLICANT: FOSS, Margit

; APPLICANT: MCNALLY, Francis J

; APPLICANT: LAURENSEN, Patricia

; APPLICANT: HERSKOWITZ, Ira

; APPLICANT: LI, Joachim J

; APPLICANT: GAVIN, Kimberly

; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT

; STREET: 4 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/484,105

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Osman Ph.D., Richard Aron

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 494-8771

; TELEFAX: (415) 494-8771

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 615 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-484-105-6

Query Match 4.8%; Score 90.5; DB 1; Length 615;  
Best Local Similarity 20.7%; Pred. No. 0.31;  
Matches 64; Conservative 54; Mismatches 110; Indels 81; Gaps 16;

QY 24 LSLFTQNSQISQRAYSTPYPS-----YVRIILQKNEKEQALARKHCISILEFPKMLF 76

Db 326 MSYFFQNAFVS---FIDFVNVDLNDLKLKIL-----SRCPTFMFVEVGLK 369

QY 77 VH-----LILSKNOREGCS---TDMAVVSTP-----PFNENLWYRLSSRFSMLKSYC 122

Db 370 QHAPADEILSLITNKNRGLEFFVEFLVRENPIHGAKFVARELEELNITNFNLIELY- 428  
QY 123 PREFLOYLEAFGLLSDFLDHOAVIKFFELETHFSYYPV-----SGFVAPH-- 167  
Db 429 -----HNLICKLSYLDKWSACKKEYKDLRHF--PIDTFQELFTLDRSGLLTQSIF 480  
QY 168 -OYLSLLQORYFPPIASVMTLDKNF-SLTPDL---IHDLLGHVPWLLHP-----SFSEF 217  
Db 481 PSYKSNIEDNLLSWEQVLPDLKENYDTLSGDLDKINAPVLGOLFPLYREANWTINIDF 540  
QY 218 FINMGRFTK--VIEKVQALPSKKORTIOTLOS-----NLIAIVRCFWFTVESGLIENH 268  
Db 541 YIAFRETLPKEILFIRDFSNYKLLLEAETPDADFVKVALILFMOAIFAFENMGLIK-F 599  
QY 269 EGRKAYGAV 277  
Db 600 QSTKSYDLV 608

RESULT 15

US-08-484-106-6

; Sequence 6, Application US/08484106

; Patent No. 5614618

; GENERAL INFORMATION:

; APPLICANT: STILLMAN, Bruce

; APPLICANT: BELL, Stephen P

; APPLICANT: KOBAYASHI, Ryuj1

; APPLICANT: RINE, Jasper

; APPLICANT: FOSS, Margit

; APPLICANT: MCNALLY, Francis J

; APPLICANT: LAURENSEN, Patricia

; APPLICANT: HERSKOWITZ, Ira

; APPLICANT: LI, Joachim J

; APPLICANT: GAVIN, Kimberly

; TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES

; NUMBER OF SEQUENCES: 24

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: FLEHR, HOBBACH, TEST, ALBRITTON & HERBERT

; STREET: 4 Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/484,106

; FILING DATE:

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Osman Ph.D., Richard Aron

; REGISTRATION NUMBER: 36,627

; REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 494-8771

; TELEFAX: (415) 494-8771

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 615 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-484-106-6

Query Match 4.8%; Score 90.5; DB 1; Length 615;  
Best Local Similarity 20.7%; Pred. No. 0.31;  
Matches 64; Conservative 54; Mismatches 110; Indels 81; Gaps 16;







```

; INFORMATION FOR SEQ ID NO: 20:
;
; SEQUENCE CHARACTERISTICS:
;
; LENGTH: 521
;
; TYPE: amino acid
;
; TOPOLOGY: linear
;
; MOLECULE TYPE: peptide
US-08-996-338-20

```

```

Query Match      4.6%; Score 87.5; DB 3; Length 521;
Best Local Similarity 21.3%; Pred. No. 0.5;
Matches 68; Conservative 50; Mismatches 106; Indels 95; Gaps 19;

QY   9 DPXYI----LKIALKURQSLSFFONSGSLQRAYSTPYSYVRIILQKENKEQALARHC 64
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db   15 EFYLKHCSLAHEIETTKSWKYSSGQHVELNPRSSRIAL-----HDC 62
                                         :|:::|:::|:::|:::|:::|:::|:::|

QY   65 ISILEFF-----KNLFVHLISLSKNOREGCSTDMAVVS-----TPPFN-- 103
      :||::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db   63 --VFLEFVPELVNDTSYFFQMKNVTQKWNLVINRNKHSCEFTQRTSKIVEVKFKQIT 120
                                         :|:::|:::|:::|:::|:::|:::|:::|

QY   104 -RNLWYRLLSRFSLMKSYCPFFLDYLEAFGLSDFLDHQAVIKFFELETHTSYSPVSG 162
       :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db   121 CENSYYOTLVNSTLYKN-CKLLLENKK-----NPTIKNAEFEDO-GYYSYCVH 168
                                         :|:::|:::|:::|:::|:::|:::|:::|

QY   163 FYAPHOYLSLLODRYPPIASVMRTLDKNFSLPDLIHLLGHVPWLHLHFSEFFINMG 222
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db   169 FL--HH-----NGKLNIRK-----TFNIT--IVEDRSNIIVPLLGPLNHAVELG 211
                                         :|:::|:::|:::|:::|:::|:::|:::|

QY   223 RLFTKVIEKVOALPSKKQRIQTLOSNIATVRCFW-FTVESGLIEN-HGRKAYGAVLIS 280
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db   212 K---NVRLNC-SALLNEEDVI-----YMFGEENGSDPNIECKE-----MRIM 251
                                         :|:::|:::|:::|:::|:::|:::|:::|

QY   281 SPOELGFAGIDNVRVLPLE 299
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db   252 TPFGKWAH---SKVLRIE 266
      :|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

```

1  RESULT 20
2  US-08-604-333-2
3  ; Sequence 2, Application US/08604333
4  ; Patent No. 5776731
5  ; GENERAL INFORMATION:
6  ; APPLICANT: Parnet, Patricia et al.
7  ; TITLE OF INVENTION: Receptor Designated 2F1
8  ; NUMBER OF SEQUENCES: 5
9  ; CORRESPONDENCE ADDRESS:
10 ; ADDRESSEE: Kathryn A. Anderson, Immunex Corporation
11 ; STREET: 51 University Street
12 ; CITY: Seattle
13 ; STATE: WA
14 ; COUNTRY: USA
15 ; ZIP: 98101
16 ; COMPUTER READABLE FORM:
17 ; MEDIUM TYPE: Floppy disk
18 ; COMPUTER: Apple Macintosh
19 ; OPERATING SYSTEM: Apple 7.1
20 ; SOFTWARE: Microsoft Word, Version 5.1a
21 ; CURRENT APPLICATION DATA:
22 ; APPLICATION NUMBER: US/08/604.333
23 ; FILING DATE: 21-FEB-1996
24 ; CLASSIFICATION: 435
25 ; ATTORNEY/AGENT INFORMATION:
26 ; NAME: Anderson, Kathryn A.
27 ; REGISTRATION NUMBER: 32,172
28 ; REFERENCE/DOCKET NUMBER: 2619
29 ; TELECOMMUNICATION INFORMATION:
30 ; TELEPHONE: (206) 587-0430
31 ; TELEX: (206) 233-0644
32 ; TELEX: 756822
33 ; INFORMATION FOR SEQ ID NO: 2:
34 ; SEQUENCE CHARACTERISTICS:
35 ; LENGTH: 541 amino acids

```

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-604-333-2

Query Match          4.6%; Score 87.5; DB 1; Length 541;
Best Local Similarity 21.3%; Pred. No. 0.53;
Matches 68; Conservative 50; Mismatches 106; Indels 95; Gaps

QY  9  DKYI-----LKIALKLPQSLFFQNSIQRAYSTPYSYRIILOENKEKQALAHKC 64
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  34  EFYLUHGCSCSLAEHETTKWSKSGSQEIVELNPRSSRIAL-----HDC 81
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY  65  ISILEFF-----KNLLFVHLLSKNQREGCSTDMAVVS-----TPFFN-- 103
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  82  --VLEFPVPELNDGTGSYFFQMKRYTOKWLNVIIRNKHSCFTEKQVTSKIVEYKKEFQIT 139
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY  104  -RNLMYRLLSSRFLSKSYCPREFFLDYLEAFGLLSDFLDHOQAVIKFFPELETHFSYYPVSG 162
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  140  CENSYSQTLVNSTSLYKN-CKKLLNNK-----NPTIKKNAEFEDQ-GYYSQVH 187
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY  163  FVAPHOYLSLLQDRFFPIASVMRTLKDKNFSLTPLIHDLGHVPMLLHFSFSEFFINMG 222
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  188  FL--HH-----NGKLFNITK-----TFNIT--IVEDRSNIYVLLGPKLHNVAELG 230
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY  223  RLFTVKIEKVALPSKKORIQTLOSLNIAIVRCFW-FTVESGLTEN-HEGRKAYGAVLIS 280
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  231  K---NVLRLNCSALLNEEDVI-----YWMFGEENGSDPNHIEEKE-----MRIM 270
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

QY  281  SPQELGHAFIDNVRVLPLE 299
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db  271  TPEGKWAH-----SKVLRIE 285
      :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-110-618-2

Query Match 4.6%; Score 87.5; DB 3; Length 541;  
Best Local Similarity 21.3%; Pred. No. 0.53;  
Matches 68; Conservative 50; Mismatches 106; Indels 95; Gaps 19;

QY 9 DPXYI---LKALKRQSLSLFQNSQSLRAYSTPYSYRILQRENKQALARKHC 64  
DB 34 EPYLLKHCSCSLAHEIETTKSWYKSSGQEHVELNPRSSRIAL-----HDC 81  
QY 65 ISLLEF-----KNLLFVHLLSLSKNOREGCGTDMAVVS-----TPFEN-- 103  
DB 82 --VLEFPVELNDGTSYFFQMKNTQKWLNVIRRNKHSCTFERQVTSKIVEVKKFQIT 139  
QY 104 -RMWTRLSSRFLMSKSYCPFRFLDYLEAFGLSDFLDHQAVIKFTELETHSYIPVSG 162  
DB 140 CENSYOTLVNSTSLYKN-CKLLLENK-----NPTIKNAEFQD-GYTSCVH 187  
163 FVAPHQYLSLQDRYFPPIASVMRLDKNFSLTPDLIHLGLHVPWLLHPSFSEFFINMG 222  
DB 188 FL--HH-----NGKLFNITK-----TFNIT--IVEDRSNIVPLGLPKLNHVAVELG 230  
QY 223 RLFTKVIKQVQALPSKQRIOTLQSLNLIATVRCFW-FTVESGLIEN-HEGRKAYGAVLIS 280  
DB 231 K---NVLNCSALLNEEDVI-----YWMFGENGSDPNIHEEKE-----NRIM 270  
QY 281 SPOELGHAFIDNVRLPLE 299  
DB 271 TPESKWH-----SKVLRIE 285

## RESULT 22

US-08-392-625-20  
Sequence 20, Application US/08392625

Patent No. 5837485

## GENERAL INFORMATION:

APPLICANT: Entian, Karl-Dieter  
APPLICANT: G tz, Friedrich  
APPLICANT: Schnell, No. 5837485bert  
APPLICANT: Augustin, Johannes  
APPLICANT: Engelke, Gerhard  
APPLICANT: Rosenstein, Ralf  
APPLICANT: Kaletta, Cortina  
APPLICANT: Klein, Cora  
APPLICANT: Wieland, Bernd  
APPLICANT: Kupke, Thomas  
APPLICANT: Jung, G nther  
APPLICANT: Kellner, Roland

TITLE OF INVENTION: Biosynthetic Process For The Preparation

TITLE OF INVENTION: Of Chemical Compounds

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Avenue

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/392,625

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/876,791

FILING DATE: 30-APR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Esmond, Robert W.  
REGISTRATION NUMBER: 32,893  
REFERENCE/DOCKET NUMBER: 0652.0980002  
TELEPHONE: (202) 371-2600  
TELEFAX: (202) 371-2540  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 990 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-392-625-20

Query Match 4.6%; Score 86.5; DB 2; Length 990;

Best Local Similarity 23.1%; Pred. No. 1.7;

Matches 57; Conservative 38; Mismatches 79; Indels 73; Gaps 13;

QY 3 YCERLDPKYILKIAL-KLRQSLSLFQNSQSLRAYSTPYSY-----RIIL 49  
DB 601 FCPRIYKNILKATWKINSEM---FSETENWLNRFATIRKWHIPKDVIIAFGDNELL 657  
QY 50 OKENKE-----QALARKKICISILEFFKNLLFVHLLSLSKNOREGCGTDMAVVSTPFF-- 102  
DB 658 NLNDKHLIILKELKHGRIRILEP-----INESNNE-----MLEIYTPLYKK 703  
QY 103 -----NRNLWYRLSSRFLW-----KSYCPFRFLDYLEAFGL-----LSDFLDHQ 143  
DB 704 TSLKQSFETIPKRNKHNKWFHLSIPKTYODNFQDYLPLFFTELKVNFINKE 763  
QY 144 AVIKFELETHFSYYPVSGFVAPHQYLSLQ--DRYFPPIASVMRLDKNFSLTPDLI-H 200  
DB 764 FYIKKEDEDFIK-----LRLLEDEDSQIYSFKIKNW-KDYCLLNSELYDY 809  
QY 201 DLIGHVP 207  
DB 810 SIVDYVP 816

## RESULT 23

US-08-466-961A-20

Sequence 20, Application US/08466961A

Patent No. 5843709

## GENERAL INFORMATION:

APPLICANT: Entian, Karl-Dieter  
APPLICANT: G tz, Friedrich  
APPLICANT: Schnell, No. 5843709bert  
APPLICANT: Augustin, Johannes  
APPLICANT: Engelke, Gerhard  
APPLICANT: Rosenstein, Ralf  
APPLICANT: Kaletta, Cortina  
APPLICANT: Klein, Cora  
APPLICANT: Wieland, Bernd  
APPLICANT: Kupke, Thomas  
APPLICANT: Jung, G nther  
APPLICANT: Kellner, Roland

TITLE OF INVENTION: Biosynthetic Process for the Preparation of

TITLE OF INVENTION: Chemical Compounds

NUMBER OF SEQUENCES: 42

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sterne, Kessler, Goldstein & Fox

STREET: 1100 New York Avenue, NW

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20005

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,961A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/392,625
; FILING DATE: 22-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/876,791
; FILING DATE: 30-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/784,234
; FILING DATE: 31-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 0652.0980004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540 20:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 990 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-466-961A-20

Query Match 4.6%; Score 86.5; DB 2; Length 990;
Best Local Similarity 23.1%; Pred. No. 1.7;
Matches 57; Conservative 38; Mismatches 79; Indels 73; Gaps 13;

QY 3 YCERLDPKYLKIAL-KLRQSLFFQNSQSLQRAYSTPYSY-----RIIL 49
Db 601 FCPRIIYNIILKPKATWKINSEM---FSETENLNRFPATIRKWHIPKDVIIAFGDNRL 657
QY 50 OKENKE-----KQALARKKICITILEFFKLLFVHLLSLKNORECCSDMAVSTPFF-- 102
Db 658 NLNDKHLIILKELKHGRIRLESF-----INESNER-----MLEIYVPLYK 703
QY 103 -----NNLWYRLSSRSFLW---KSYCPREFLDYLEAFGL---LSDFLDHQ 143
Db 704 TSLKQSQSIIPKRNKFNKLNKOWFSIHLSIPTQDNFIQDYLPLPFTTELKVNFNK 763
QY 144 AVIKFELETHFSYYPVSGFVAPHOYLSLQ--DRYFPFIASVMRTLDKDNFSLTPDLI-H 200
Db 764 FYIKFEDEDFIK-----LRLREDEYSQIYSFIKNW-KDYCLLNSELYDY 809

QY 201 DLLGHVP 207
Db 810 SIVDVP 816

RESULT 24
US-08-645-193B-15
Sequence 15; Application US/08645193B
Patent No. 5962253
GENERAL INFORMATION:
APPLICANT: Kupke, Thomas
APPLICANT: Gotz, Friedrich
APPLICANT: Kempter, Christoph
APPLICANT: Jung, Gunther
TITLE OF INVENTION: Oxidative Decarboxylation of Peptides
TITLE OF INVENTION: Catalyzed by Flavoprotein Epd
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:

;
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/645,193B
; FILING DATE: 13-MAY-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 0652.1540000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 990 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; US-08-645-193B-15

Query Match 4.6%; Score 86.5; DB 2; Length 990;
Best Local Similarity 21.7%; Pred. No. 1.7;
Matches 62; Conservative 43; Mismatches 100; Indels 81; Gaps 14;

QY 3 YCERLDPKYLKIAL-KLRQSLFFQNSQSLQRAYSTPYSY-----PYSYRII 48
Db 601 FCPRIIYNIILKPKATWKINSEM---FSETENLNRFPATIRKWHIPKDVIIAFGDNRL 657
QY 49 OKENKE-----KQALARKKICITILEFFKLLFVHLLSLKNORECCSDMAVSTPFF-- 102
Db 658 NLNDKHLIILKELKHGRIRLESF-----INESNER-----MLEIYVPLYK 703
QY 103 -----NNLWYRLSSRSFLW---KSYCPREFLDYLEAFGL---LSDFLDH 142
Db 704 TSLKQSQSIIPKRNKFNKLNKOWFSIHLSIPTQDNFIQDYLPLPFTTELKVNFNK 763
QY 143 QAVIKFELETHFSYYPVSGFVAPHOYLSLQ--DRYFPFIASVMRTLDKDNFSLTPDLI- 199
Db 764 FYIKFEDEDFIK-----LRLREDEYSQIYSFIKNW-KDYCLLNSELYD 809

QY 200 DLLGHVP 207
Db 810 SIVDVP 816

RESULT 25
5194600-4
Patent No. 5194600
APPLICANT: BUSSEY, HOWARD;BOONE, CHARLES;SOMMER, STEVE S.;
HILL, KATHRYN;MEADEN, PHILIP
TITLE OF INVENTION: GENES WHICH PARTICIPATE IN B-GLUCAN
ASSEMBLY AND USE THEREOF
NUMBER OF SEQUENCES: 4
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,316
FILING DATE: 03-MAR-1990
SEQ ID NO: 4
LENGTH: 1365
5194600-4

Query Match 4.4%; Score 83.5; DB 6; Length 1365;
Best Local Similarity 19.0%; Pred. No. 5.9;
Matches 74; Conservative 49; Mismatches 100; Indels 167; Gaps 19;

QY 89 GCSTDMVAVSTPFF--FNRNLWYRLSS-----RFSLMKSYC----- 122
Db 172 GCPT---VIDSDFEEFNRLPHEAMNGEGKFRF--IWRSTCSLDGKSVYPLTHPLEITLQ 227
```

```
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/466,961A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/392,625
; FILING DATE: 22-FEB-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/876,791
; FILING DATE: 30-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/784,234
; FILING DATE: 31-OCT-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Esmond, Robert W.
; REGISTRATION NUMBER: 32,893
; REFERENCE/DOCKET NUMBER: 0652.0980004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540 20:
; INFORMATION FOR SEQ ID NO: 20:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 990 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-466-961A-20

Query Match 4.6%; Score 86.5; DB 2; Length 990;
Best Local Similarity 23.1%; Pred. No. 1.7;
Matches 57; Conservative 38; Mismatches 79; Indels 73; Gaps 13;

QY 3 YCERLDPKYLKIAL-KLRQSLFFQNSQSLQRAYSTPYSY-----RIIL 49
Db 601 FCPRIIYNIILKPKATWKINSEM---FSETENLNRFPATIRKWHIPKDVIIAFGDNRL 657
QY 50 OKENKE-----KQALARKKICITILEFFKLLFVHLLSLKNORECCSDMAVSTPFF-- 102
Db 658 NLNDKHLIILKELKHGRIRLESF-----INESNER-----MLEIYVPLYK 703
QY 103 -----NNLWYRLSSRSFLW---KSYCPREFLDYLEAFGL---LSDFLDHQ 143
Db 704 TSLKQSQSIIPKRNKFNKLNKOWFSIHLSIPTQDNFIQDYLPLPFTTELKVNFNK 763
QY 144 AVIKFELETHFSYYPVSGFVAPHOYLSLQ--DRYFPFIASVMRTLDKDNFSLTPDLI-H 200
Db 764 FYIKFEDEDFIK-----LRLREDEYSQIYSFIKNW-KDYCLLNSELYDY 809

QY 201 DLLGHVP 207
Db 810 SIVDVP 816

RESULT 24
US-08-645-193B-15
Sequence 15; Application US/08645193B
Patent No. 5962253
GENERAL INFORMATION:
APPLICANT: Kupke, Thomas
APPLICANT: Gotz, Friedrich
APPLICANT: Kempter, Christoph
APPLICANT: Jung, Gunther
TITLE OF INVENTION: Oxidative Decarboxylation of Peptides
TITLE OF INVENTION: Catalyzed by Flavoprotein Epd
NUMBER OF SEQUENCES: 70
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, Suite 600
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005
COMPUTER READABLE FORM:
```

us-09-438-185-1047.ra1

Mon Jul 2 08:46:40 2001

QY 123 -----PREFLD-----YLEAFGLLSDFLDH--- 142  
Db 228 NGRMSSIPOLKKILYTPKREILVGADNKKQLHDLPEPELREKLKRVTSLSISEFYQYKKD 287  
QY 143 -----QAVIKFF-----ELETHFSYYPVSGF----- 163  
Db 288 ITATLNFTRKSIIVNPEPLISKOLIKVSSVKNKDIITSNEELNSKGFDPYNNMLGLYINGQNWKI 347  
QY 164 --VAPHOYLSLQDRYFPPIASVMRTLDK-----DNFSLTPDLIHDLLGHVPWL--- 209  
Db 348 TSLTPYNLLTALKTEYQSLLKITNLQLELEPSKCILDSKFLLNKFSQFSLGKLNLOPIK 407  
QY 210 --LH--PSSEFFINMGR-----FTKVIEKVOAL--PSKKORIOTLOSNIATVRCFWF 258  
Db 408 MDLHTIPGSESVIYFNDIESDPQYDELVNSVQAFFDKSKFGELPEIKONWSEII---F 463  
QY 259 TVESGLIENHEGRKAYG-----AVLISS---POELGHAFIDNVRLPLEDQIIRLPENTS 311  
Db 464 VIDFARLESEVKEALGGLVRVNVVSGTQPRVG-----LLPFSSD 505  
312 TPQETLFSINHF-----DELVELTSKLEWML 337  
506 SDKSVNKKIYELKNSTDNI/TELKSFLETML 535

Search completed: June 29, 2001, 13:56:34  
Job time: 223 sec

**THIS PAGE BLANK (USPTO)**









```
100 .....ProPhePhe..... 102
556 TCAAGTCAAAAAGAGCCCTTCTTCTGGTAACAAAAACACTGATATGCT 605
102 ..... 102
606 ATTGCACAAATATTATATTGTTTCCAGCAGCAAAATCAATGATAAATGAGC 655
103 ..... 103
656 AGAATACATCATCAATCATATATAAGAAAAACACACAGATAAAGTTGAAT 705
110 LeuLeuSerSerArgPheSerLeu..... 117
706 ATTCTCAAAATATTATCTAGTTTAATTTTGAATAATTTTCAAGACAATC 755
118 ..... 118
756 TTTAGAGGTGTTTATATTCCACATCACTAATGGGAAATTTACAGCCCT 805
124 rgPhePhe..... 126
806 TACTTTTGAATAGTCTCAACGCCGACACTAATCAAAATTTCTGATATG 855
127 LeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheLeuAspHisG1 143
856 AAGAAAAAATTAGAGTTAATAGTTAGTACAGAAACGAGTCATTGAAGAATAA 905
143 nAlaValIleLysPhePheGluLeuGluThrHisPheSerTyrTyrProV 160
906 TTCTTCTATACGA..... 928
160 alSerGlyPheValAlaProHisGlnTyrLeuSerLeuLeuGlnAspArg 176
929 TGCTGCCAGCAACCAATGACATTTCAATTAATCTATCGTACCTGTATAA 978
177 TyrPheProIleAlaSerValMet..... 184
979 CAATTTCCAAAACACTCATGATCAATCAATTAATGATATAGTACTCAGTT 1028
185 ....ArgThrLeuAspLysAspAsn..... 191
1029 CACCAAGCTACTGGAC...GATAACCTCAGCAATTCGATTGGCCCACTT 1075
192 ..PheSerLeuThrProAspLeuIleHisAspLeuLeuGlyHisValPro 207
076 AC GTTGACCAAAATCCCTTACAATGCATAAAATTTATCAATTAATTTCTC 1125
208 TrpLeuLeuHisProSer.....PheSerGluPhePheIleAsnMetG1 222
1126 TGGTCCATACATCCATCAAGCAATTTGATCACTATCAATCTAAT..... 1170
222 YArgLeuPheThrLysValIleGluLysValGlnAlaLeuProSerLysL 239
1171 .....CAACTGGTAGGCAATTTACTATTG..... 1197
239 ysglnArgIleGlnThrLeuGlnSerAsnLeu.....IleAlaIleVal 253
1198 .....CGAATAAATTCACAGATGAGGATTTGCACGAATTCAGATAGAA 1242
254 ArgCysPheTrpPheThrValGluSerGlyLeuIleGluAsnHisGluG1 270
1243 GATGCCATTTGGTTCATTTGTTTCCAA...TTAGCCAAAAAATTTTTCGC 1289
270 YArgLys..... 272
1290 CCAAAAGAGGTGGTATCATATATATGTCCTTCTTTGTATCGCTGCTTA 1339
273 .....AlaTyrGlyAlaVal..... 277
1340 ATATACTAATTACTATGTCATCATTAAGTCCCTACGTATATCAGAAG 1389
278 LeuIleSerSerPro.....GlnGluLeuGlyHisAlaPheI1 290
```

```
1190 CTRATCAGTTCCGGCCTACTTTATCTCCAAGATTCTCAATGATAAGTTTGT 1439
290 eAsp.....AsnValArgValLeuProLeuGluLeuAspG 302
1440 GCATGTCAGCTGTAAATTAACTTGAAAATTTACCGTGTGATGAAAAGTC 1489
302 InIleIleArgLeuProPheAsnThrSerThrProGlnGluThrLeuPhe 318
1490 AATACATATGTTATTTGAGGAACGTTATGGAATATGACGTTAAATTTAT 1539
319 SerIleArgHisPheAspGluLeuValGluLeuThrSerLysLeu..... 333
1540 GAAATTTTAAATTTTCGCCCACTCGTGAATATCACAGAACAAATCAAAAT 1589
333 ..... 333
1590 GCGAATACTCTCCAATGATATAACTAAATTTGCAACTGTCGAAAACCTCTC 1639
334 .....GluTrpMetLeuAsp.....Gln 339
1640 TGACATPATAAATCATGTTTGCAGAAATGTTACTTATCACATTTATGTTCC 1689
340 GlyLeuLeuGluSerIleProLeuTyrAsnGlnGluLysTyrLeuSerG1 356
1690 GGTATTTTATCTAGTGT.....AACGCCACAGTGTGCTAAAAAT 1730
356 yPheGluValLeuCys 361
1731 ATTCAGATTTTGT 1746
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-218-265-11
seq_documentation_block:
; Sequence 11, Application US/08218265
; Patent No. 5922585
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; TITLE OF INVENTION: NO. 5922585el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Thereof
; NUMBER OF SEQUENCES: 35
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: US
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/218,265
; FILING DATE: 25-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WH194-03
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
```

us-09-438-185-1047.rni

Mon Jul 2 08:46:43 2001

FEATURE: ;  
 NAME/KEY: CDS ;  
 LOCATION: 241..3918 ;  
 US-08-218-265-11

alignment\_scores:      Quality: 95.00      Length: 522  
                          Ratio: 0.448      Gaps: 25  
                          Percent Similarity: 18.774

alignment\_block:

US-09-438-185-1047 x US-08-218-265-11 ..

Align seg 1/1 to: US-08-218-265-11 from: 1 to: 4002

```

12 TyrLeuLysIleAlaLeuLysLeuArgGlnSerLeuSerLeuPhePhe 28
|||||
277 TACATCTGGAA.....AAGCTCATCTTTGATGACAAATCACTA 317
|||||
28 ecInAsnSerGlnSerLeuGlnArgAlaTyrSerThrProTyrSerTyr 45
|||||
318 TAAGGATTCTCAACAACG...CGTACGTGGANGAGCGAGATTCTTATT 364
|||||
45 yArgIleIle.....LeuGlnLysGluAsnLysGlu 55
|||||
365 TTTTAAACATTTTGGGAATTGCTACTCACTAAGATTGATCAATAAGGAA 414
|||||
56 LysGlnAlaLeuAlaArgHisLysCysIleSerIleLeuGluPhePhe 72
|||||
415 .....ATCITTATCATCTGGCTTGTAGAGTTTATAATAAGATGGA 455
|||||
72 sasLeuLeuPhe.....ValHisLeuSerLeuSerLysA 85
|||||
456 AAACCTCGAATTTTCCCATATTCITTACATATTTGTATGATTTTTTGA 505
|||||
85 sngInArgGluGlyCysSerThrAspMetAlaValSerThr..... 99
|||||
506 AGCAGATCTGCCAATTTGATACAAATGCTCTCTGTTCGGGTACAAATA 555
|||||
100 .....ProPhePhe..... 102
|||||
556 TCAAGTCAAAAGAGCCCTCTTCTTCTGTGTAACMAAAATCACTGATGCT 605
|||||
102 ..... 102
606 ATTGCACAAAATATTATATTGTTTCCAGCAGCAAAATCAATGATAATGACG 655
|||||
103 .....AsnArgAsnLeuTyrTyrArg 109
|||||
656 AGAACTACATCATCAATGATATTAAGAAACAAACAAAGATAAAGTTGAA 705
|||||
110 LeuLeuSerSerArgPheSerLeu..... 117
|||||
706 ATTCACAAATATTATCCAGTTTAATTTTGAAATTTTCAAGAACATC 755
|||||
118 .....TrpLysSerTyrCysPro 124
|||||
756 TTPAGAGGTGTTTATATTTCACATCTAATCTGGGAAATTTTACAGCCCT 805
|||||
124 rgPhePhe..... 126
|||||
806 TACTTTTGAATAGTCTCAACCCGACACTAATCAAAATCTCATATG 855
|||||
127 LeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheLeuAspHisG 143
|||||
856 AAGAAAAAATTAGAGTTAATTAGTAGTACGAACGAGCTCAITTCAGAA 905
|||||
143 nAlaValIleLysPhePheGluLeuGluThrHisPheSerTyrTyrPro 160
|||||
906 TTCTTCTATACGA.....AACGTAATAA 928
|||||
160 alSerGlyPheValAlaProHisGlnTyrLeuSerLeuLeuGlnAspArg 176

```

```

929 TGTCTGCCAGCAACGCAAAATGACTTTTCAATTAAGTATCGTCACCTGATAA 978
|||||
177 TyrPheProIleAlaSerValMet..... 184
|||||
979 CAAATTTCCAAACATATCATGCAATTAATTAATTTGTATAGATACTCACTT 1028
|||||
185 ....ArgThrLeuAspLysAspAsn..... 191
|||||
1029 CACCAAGCTACTGGAC...GATAACCCCTACAGAAATTCGATTGGCCCACTT 1075
|||||
192 ..PheSerLeuThrProAspLeuIleHisAspLeuLeuGlyHisValPro 207
|||||
1076 ACCTTGACCAAAATCCCTTACAAATGCATAAAATTTATCAATTAATCTC 1125
|||||
208 TrpLeuLeuHisProSer.....PheSerGluPhePheIleAsnMetG 222
|||||
1126 TGTCCATATCATCCATCAAGCAATTTGATCACTATCAATCTAAT..... 1170
|||||
222 yArgLeuPheThrLysValIleGluLysValGlnAlaLeuProSerLysL 239
|||||
1171 .....CAACTGGTAGCGAAATTTACTATTG..... 1197
|||||
239 ysGlnArgIleGlnThrLeuGlnSerAsnLeu.....IleAlaIleVal 253
|||||
1198 ....CGATAAATTCACAGATGAGGATTTGACGAATTCAGATAGAA 1242
|||||
254 ArgCysPheThrPheThrValGluSerGlyLeuIleGluAsnHisGluG 270
|||||
1243 GATGCCATTTGGTCATGTTTTCCTCA...TTACCCAAAATTTTCGCG 1289
|||||
270 yArgLys..... 272
|||||
1290 CAAAAGAGGGTGGTATCATATATATGATCCCTCTTTGTATCGCTGCTTA 1339
|||||
273 .....AlaTyrGlyAlaVal..... 277
|||||
1340 ATATACTAATTAATCTATGTCATCAATTAAGTCCCTACGTATATACAGAA 1389
|||||
278 LeuIleSerSerPro.....GlnGluLeuGlyHisAlaPheI 290
|||||
1390 CTAATCAGTTCCGCGCTACTTTATCTCCAAGATTCCAATGATAAGTTGT 1439
|||||
290 easp.....AsnValArgValLeuProLeuGluLeuAspG 302
|||||
1440 GCATGTCAGCTGTTAATTAACCTGAAATTTACCCGTTGATGAAAAGTC 1489
|||||
302 InIleIleArgLeuProPheAsnThrSerThrProGlnGluThrLeuPhe 318
|||||
1490 AATACATATGTTATGAGCAAGTTTATGGAATATGACGTTAAATTTTAT 1539
|||||
319 SerIleArgHisPheAspGluLeuValGluLeuThrSerLysLeu..... 333
|||||
1540 GAATTTTAAATTCGACCAACCTCGTGGAAATTCACAGAACAAATCAAAAT 1589
|||||
333 ..... 333
1590 GCGAATACTCTCAATCATATATACTAATTTGCAACTGTCGAAACTCCTC 1639
|||||
334 .....GluTrpMetLeuAsp.....Gln 339
|||||
1640 TGAGCATTAATAATCATGGTTGCAAGATGTTACTTATCACTATATGTTCC 1689
|||||
340 GlyLeuLeuGluSerIleProLeuTyrAsnGlnGluLysTyrLeuSerG 356
|||||
1690 GGTATTTTATCTAGTGT.....AACCGCACAGTGTGCTAAAAAT 1730
|||||
356 yPheGluValLeuCys 361
|||||
1731 ATTCAGATTTTGT 1746

```

seq\_name: /cgn2\_6/ptodata/2/1na/6A\_COMB.seq:US-08-521-872-11

seq\_documentation\_block:  
; Sequence 11, Application US/08521872  
; Patent No. 6015682  
; GENERAL INFORMATION:  
; APPLICANT: Young, Richard A.  
; APPLICANT: Koleske, Anthony J.  
; APPLICANT: Thompson, Craig M.  
; APPLICANT: Chao, David M.  
; TITLE OF INVENTION: No. 6015682el Factors Which Modify Gene  
; TITLE OF INVENTION: Transcription and Methods of Use Therefor  
; NUMBER OF SEQUENCES: 37  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
; STREET: Two Militia Drive  
; CITY: Lexington  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02173

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/521,872  
; FILING DATE: 31-AUG-1995  
; CLASSIFICATION: 436  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/218,265  
; FILING DATE: 25-MAR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Granahan, Patricia  
; REGISTRATION NUMBER: 32,227  
; REFERENCE/DOCKET NUMBER: WH194-03A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-861-6240  
; TELEFAX: 617-861-9540  
; INFORMATION FOR SEQ ID NO: 11:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4002 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 241..3918  
; PUBLICATION INFORMATION:  
; DOCUMENT NUMBER: US 08/218,265  
; FILING DATE: 25-MAR-1994  
; US-08-521-872-11

alignment\_scores:  
Quality: 95.00 Length: 522  
Ratio: 0.448 Gaps: 25  
Percent Similarity: 40.613 Percent Identity: 18.774

alignment\_block:  
US-09-438-185-1047 x US-08-521-872-11 ..  
Align seq 1/1 to: US-08-521-872-11 from: 1 to: 4002  
12 TyrTleLeuLysIleAlaLeuLysLeuArgGlnSerLeuSerLeuPhePhe 28  
|||||  
277 TACATCCTGGAA.....AAGCTCATCTTTGATATGACAAATCACTA 317  
28 eGlnAnSerGlnSerLeuGlnArgAlaTyrSerThrProTyrSerTyr 45  
|||||  
318 TAACGATCTCAACAACG...CGTACGTGAAGAGGACGATTTCTTATT 364  
45 yArgIleIle.....LeuGlnLysGluAsnLysGlu 55  
|||||

365 TTTTAAACACTTTTGGGGAATTCCTACTCACTAAGATTGATCAATAAGGAA 414  
56 LysGlnAlaLeuAlaArgHisLysCysIleSerIleLeuGluPhePhe 72  
::: |||||  
415 .....ATCTTTTCATCATGGCTTGTAGAGTTTATAAATAAGATGGA 455  
72 sAsnLeuLeuPhe.....ValHisLeuLeuSerLeuSerLysA 85  
|||||  
456 AAACCTTCGATTTTGGCCATTATCTTTACATATTTTTCATGATTTTGG 505  
85 sGlnArgGluGlyCysSerThrAspMetAlaValSerThr..... 99  
|||||  
506 ACACATCTGCCAAATGATACAAATGCTCTCTGTGGCTACATAACA 555  
100 .....ProPhePhe..... 102  
|||||  
556 TCAAGTCAAAAAGAGCCCTTCTTCTGTTAAACAAAATCACTGATATGCT 605  
102 ..... 102  
606 ATTGCACAAATATTATTTGTTTCCAGCAGCAAAATCAATGATAATGAGC 655  
103 .....AsnArgAsnLeuTyrTyrArg 109  
|||||  
656 AGAATACATCATCAATGATATAAGAAAAACAACAAGATAAAGTTGAAT 705  
110 LeuLeuSerSerArgPheSerLeu..... 117  
|||||  
706 ATTCTCAAAATATTATCCAGTTTAAATTTTGAATAATTTTCAAGACAATC 755  
118 .....TyrPheSerTyrCysProA 124  
756 TTAGAGGTGTTTATATTTCCACATCTAATCTGGGAAATTTACAGCCCT 805  
124 rgPhePhe..... 126  
806 TACTTTTGAATAGTCTCAACGCCGACATAATCAAAATTCGTATG 855  
127 LeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheLeuAspHis 143  
|||||  
856 AAGAAAAATTTAGAGTTAATTATTTACAGAAACAGTCATTTGAAGATAA 905  
143 nAlaValIleLysPhePheGluLeuGluThrHisPheSerTyrTyrPro 160  
|||||  
906 TTCTTCTATACGA.....AAGTAAATAA 928  
160 alSerGlyPheValAlaProHisGlnTyrLeuSerLeuGlnAspArg 176  
|||||  
929 TGCTGCCAGCAACCAATGACTTTTCAATTAACTATCGTCACCTGTAATA 978  
177 TyrPheProIleAlaSerValMet..... 184  
979 CAATTTCCAAAACACTATCATGCATTCATTAATTTGTATAGATACCTCAGTT 1028  
185 .....ArgThrLeuAspLysAspAsn..... 191  
1029 CACCAAGCTACTGGAC...GATAACCTTACAGAAATTCGATGGCCACTT 1075  
192 ..PheSerLeuThrProAspLeuIleHisAspLeuLeuGlyHisValPro 207  
||| :|||  
1076 ACCTTGACCAAAATCCCTTACATGTCATAAAATTTCAATTAATTCCTC 1125  
208 TrpLeuLeuHisProSer.....PheSerGluPhePheIleAsnMetG 222  
||| :|||  
1126 TGTTCATATACATCCATCAAGCAATTTGATCCTATGAATCTAAT..... 1170  
222 yArgLeuPheThrLysValIleGluLysValGlnAlaLeuProSerLysL 239  
|||||  
1171 .....CAACTGGTAGCGAAATTTACTATTG..... 1197  
239 ysGlnArgIleGlnThrLeuGlnSerAsnLeu.....IleAlaIleVal 253  
|||||  
1198 .....CGAATAATTCACAGATGAGGATTTGCACGAATTCACAGATAGA 1242

us-09-438-185-1047.ini

Mon Jul 2 08:46:43 2001

```
254 ArgCysPheThrPheThrValGluSerGlyLeuIleGluAsnHisGluG1 270
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
1243 GATGCCATTGGTCATCTGGTTCCAA...TTAGCCAAAATTTTTCGGC 1289
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
270 YATGLYS..... 272
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
1290 CCAAGAGGGTGGTATATATATATATATATATATATATATATATATAT 1339
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
273 .....AlaTrpGlyAlaVal..... 277
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
1340 ATATACTATTACTATGTCATTAAGTCCCTACGTATATATATATATATAT 1389
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
278 LeuIleSerSerPro.....GlnGluGlyHisAlaPheI1 290
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
1390 CTAATCAGTTCCGGCTACTTATCTCCAGATCCCAATGATAAGTTTGT 1439
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
290 eASP.....AsnValArgValLeuProLeuGluLeuASP 302
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
1440 GCATGCCAGCTGTTAATTAATCTGAAATTTTCCGTTGATCAAAAGTC 1489
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
302 InIleIleArgLeuProPheAsnThrSerThrProGlnGluThrLeuPhe 318
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
1490 AATCAATATGTTGATGAGAACGTTATGGAATATGACGTTAAATTTTAT 1539
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
319 SerIleArgHisPheASPLeuValGluLeuThrSerLysLeu..... 333
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
1540 GAAATTTTAAATTCGACCAACTCGTGGAAATTCACAGAACAAATCAAAAT 1589
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
333 ..... 333
1590 GCGAATCTCTCCAAATGATATACTAATTTGCAACTGTGAAAACCTCCTC 1639
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
334 .....GluTrpMetLeuASP.....Gln 339
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
1640 TCAGCATTAATCAATCAAGTTGTCAGATGTGACTATATCATCTTTATGTC 1689
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
340 GlyLeuLeuGluSerIleProLeuArgAsnGlnGluLysTrpLeuSerG1 356
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
1690 GGTATTTTATCTAGTGT.....AACCCACAGTGTGTCTTAAAT 1730
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
356 yPheGluValLeuCys 361
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
1731 ATTCAGATTTTGT 1746
      ::::::::::: ||| ::| ::| ::| ::| ::| ::| ::| ::|
```

seq\_name: /cgn2\_6/ptodata/2/lna/6B\_COMB.seq:US-08-590-399-11

```
seq_documentation_block:
; Sequence 11, Application US/08590399
; Patent No. 6214588
; GENERAL INFORMATION:
; APPLICANT: Young, Richard A.
; APPLICANT: Koleske, Anthony J.
; APPLICANT: Thompson, Craig M.
; APPLICANT: Thompson, David M.
; APPLICANT: Chao, David M.
; TITLE OF INVENTION: No. 6214588el Factors Which Modify Gene
; TITLE OF INVENTION: Transcription and Methods of Use Therefor
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Willitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/590,399
; FILING DATE: 26-JAN-1996
```

```
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/540,804
; FILING DATE: 11-OCT-1995
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/521,872
; FILING DATE: 31-AUG-1995
PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/218,265
; FILING DATE: 25-MAR-1994
ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: WHI94-03A3
TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4002 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 241..3918
; US-08-590-399-11

alignment_scores:
Quality: 95.00 Length: 522
Ratio: 0.448 Gaps: 25
Percent Similarity: 40.613 Percent Identity: 18.774

alignment_block:
US-09-438-185-1047 x US-08-590-399-11 ..
Align seg 1/1 to: US-08-590-399-11 from: 1 to: 4002
12 TyrIleLeuLysIleAlaLeuLysLeuArgGlnSerLeuSerLeuPhePh 28
||||| ::| ::| ::| ::| ::| ::| ::| ::|
277 TACATCTCTGAA.....AAGCTCATCTTTGATGATGACAAATCACTA 317
||||| ::| ::| ::| ::| ::| ::| ::| ::|
28 eGlnAsnSerGlnSerLeuGlnArgAlaTrpSerThrProTrpSerTrp 45
||||| ::| ::| ::| ::| ::| ::| ::| ::|
318 TACGATTCCTCAAACTG...CGTACGTGGAAGAGGACAGATTCTTATT 364
||||| ::| ::| ::| ::| ::| ::| ::| ::|
45 yArgIleIle.....LeuGlnLysGluAsnLysGlu 55
||||| ::| ::| ::| ::| ::| ::| ::| ::|
365 TTTTAAACCTTTGGGGAATTGCTACTCACTAAGATTGATCAATAAGGAA 414
||||| ::| ::| ::| ::| ::| ::| ::| ::|
56 LysGlnAlaLeuAlaArgHisLysCysIleSerIleLeuGluPhePhe 72
||||| ::| ::| ::| ::| ::| ::| ::| ::|
415 .....ATCTTCATCATCTGGCTTGTAGAGTTTATAAATAAGATGGA 455
||||| ::| ::| ::| ::| ::| ::| ::| ::|
72 AsnLeuLeuPhe.....ValHisLeuLeuSerLeuSerLysA 85
||||| ::| ::| ::| ::| ::| ::| ::| ::|
456 AAACCTCGAATTTTGGCATTATCTTTACATATTTTGATGATTTTGGGA 505
||||| ::| ::| ::| ::| ::| ::| ::| ::|
85 snGlnArgGluGlyCysSerThrAspMetAlaValSerThr..... 99
||||| ::| ::| ::| ::| ::| ::| ::| ::|
506 ACGACATCTGCCAAATTTGATACAAATGCTCCTGTTGGCGGTACAAATCA 555
||||| ::| ::| ::| ::| ::| ::| ::| ::|
100 .....Prophe..... 102
||||| ::| ::| ::| ::| ::| ::| ::| ::|
556 TCAAGTCAAAAAGAGCCCTTCTTTCTGGTAACAAAATCACTGATATGCT 605
||||| ::| ::| ::| ::| ::| ::| ::| ::|
102 ..... 102
606 ATTGCACAAATATTATATTTGTTCCAGCAGCAAAATCAATGATAAATGACG 655
||||| ::| ::| ::| ::| ::| ::| ::| ::|
103 .....AsnArgAsnLeuTrpTrpArg 109
||||| ::| ::| ::| ::| ::| ::| ::| ::|
```

```
656 AGAACTACATCATCAATGATATATAAGAAAAACAAGATAAAGTTGAAT 705
|||||:
110 LeuLeuSerSerArgPheSerLeu..... 117
:||||:
706 ATTCTCAAAATATTATCCAGTTTAAATTTTGAAATTTTTCACAAACAATC 755
118 .....
756 TTTAGAGGTGTTTATATTTCCACATCTAATCGGAAATTTTACAAGCCCT 805
|||||:
124 rPhePhe..... 126
:||||:
806 TACTTTTGAATAGTCTCAACGCCGACACTAATCAAAATTTCTGATG 855
127 LeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheLeuAspHisG1 143
|||||:
856 AGAAAAAATTAGAGTTTAAATTTAGTTAGTACAGAACGAGTCATTGAAGAATAA 905
143 nAlaValIleLysPheGluLeuGluThrHisPheSerTyrTyrProV 160
:||||:
906 TTTCTCTATACGA..... 928
|||||:
160 a1SerGlyPheValAlaProHisGlnTyrLeuSerLeuLeuGlnAspArg 176
:||||:
929 TGTCTGCCAGCAACCAATGACATTTCAATTAATCACTATCGTCACCTGTAAA 978
177 TyrPheProIleAlaSerValMet..... 184
|||||:
979 CAATTTCCAAACTATCATGATTCATTAATTAATTTATAGATACTCAGTT 1028
185 .....ArgThrLeuAspLysAspAsn..... 191
1029 CACCAAGTACTGGAC...GATACCTCAGCAATTCGATTCGCCCACTT 1075
192 ..PheSerLeuThrProAspLeuIleHisAspLeuLeuGlyHisValPro 207
:||||:
1076 ACCTTGACCAAAATCCCTTCAATGATCAATTAATTTCAATTAATTTCTC 1125
208 TrpLeuLeuHisProSer.....PheSerGluPhePheIleAsnMetG1 222
|||||:
1126 TGTCTCATATCATCCATCAAGCAATTTGATCACTATGAATCTAAT..... 1170
222 yArgLeuPheThrLysValIleGluLysValGlnAlaLeuProSerLysL 239
:||||:
171 .....CAACTGGTAGCGAAATTTACTATTG..... 1197
239 yGlnArgIleGlnThrLeuGlnSerAsnLeu.....IleAlaIleVal 253
|||||:
1198 .....CGAATAAATTCACAGATCAGGATTTGCACCAATTCAGATAGAA 1242
254 ArgCysPheThrPheThrValGluSerGlyLeuIleGluAsnHisGluG1 270
:||||:
1243 GATGCCCTTTGTCATTCGTTTCCAA...TTAGCCCAAAATTTTTCGGC 1289
270 yArgLys..... 272
:||||:
1290 CCAAAAGAGGGTGGTATCATATATGATGCTTCTTTGATCGCCCTGCTTA 1339
273 .....AlaTyrGlyAlaVal..... 277
1340 ATATACTAATTAATTTATGTCATCATTAAGTCCCTACGTATATCAGAAAG 1389
278 LeuIleSerSerPro.....GlnGluLeuGlyHisAlaPheI1 290
|||||:
1390 CTAAATCAGTTCCGGCTACTTTATCTCCAAGATTCCAATGATAAGTTGT 1439
290 eAsp.....AsnValArgValLeuProLeuGluLeuAspG 302
:||||:
1440 GCATGTCACGCTGTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1489
302 InIleIleArgSerProPheAsnThrSerThrProGlnGluThrLeuPhe 318
|||||:
1190-438-185-1047 x US-08-559-260-1/rev. ..
```

```
1490 AATACAATATGCTATGAGGAACGTTATGGAATATGACGTAAATTTTAT 1539
319 SerIleArgHisPheAspLeuLeuValGluLeuThrSerLysLeu..... 333
:||||:
1540 GAAATTTTAAATTTGACCAACTGCTGGAATCAGACAGACAAATCAAAAT 1589
333 ..... 333
1590 GCGAATACCTCCTCAATGATATAAATAATTTGCAACTGTCGAAACCTCTC 1639
334 .....GluTrpMetLeuAsp.....Gln 339
1640 TGAGCATTAATAATCATGTTGTCAGAAATGGTACTTATCACAATTTATGTTCC 1689
340 GlyLeuLeuGluSerIleProLeuTyrAsnGlnGluLysTyrLeuSerG1 356
|||||:
1590 GGATTTTATCTAGTGT.....AACGGCACAGTGTGCTAAAAAT 1730
356 yPheGluValLeuCys 361
|||||:
1731 ATTCAAGATTTTGT 1746
seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-559-260-1
seq_documentation_block:
; Sequence 1, Application US/08559260
; Patent No. 5656497
; GENERAL INFORMATION:
; APPLICANT: Zeikus, Joseph G.
; TITLE OF INVENTION: HYPERTHERMOSTABLE XYLOSE ISOMERASE AND
; TITLE OF INVENTION: ITS PRODUCTION
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: WI
; COUNTRY: USA
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/559,260
; FILING DATE: 15-NOV-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/302,131
; FILING DATE: 06-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Kryshak, Thad F.
; REGISTRATION NUMBER: 19,428
; REFERENCE/DOCKET NUMBER: 66-336-9045-4
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2522 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-559-260-1
alignment_scores:
Quality: 94.00 Length: 418
Ratio: 0.556 Gaps: 22
Percent Similarity: 40.431 Percent Identity: 19.139
alignment_block:
US-09-438-185-1047 x US-08-559-260-1/rev. ..
```











11132 GTTCCTCAAGTAATTCGGCAGGACCAAGACACTGATCATCCGAGAA 11181  
315 ..... 315  
11182 TGGAGCTTACGACAGACTCAGTGCATTTATCAGCAGTGTCTCAAGAAG 11231  
316 ..... 316  
11232 TACTGCCTTAATTCGAGATATGAGACCATCAGCTGTTGTCACAGAGG.. 11279  
322 sphaSpGluLeuValGluLeuThrSerLysLeuGluTrpMet..... 336  
11280 .CTAATGAGATTTACGATGTCCTCATTTTCCAGTGGTGCATRAA 11328  
337 ..LeuAspGlnGlyLeuLeu..... 342  
11329 GGTTCGAGACCTCTGCTCTGTATGTAAAGTACCCCTCATGCCCCCGAC 11378  
343 .....GluSerIleProLeuTyr.....AsnGlnGlu..... 351  
11379 CTTCAGGCCCATATCCCGTTATATAAGTCCCAATGATCAAAATCTTCAT 11428  
352 .LysTyr...LeuSerGlyPheGluValLeuCysGln 362  
11429 TAAGTACCTATGGAGGTATAGAGGATTTGTCAG 11465

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:US-08-905-817-1

seq\_documentation\_block:  
: Sequence 1, Application US/08905817  
: Patent No. 5824777  
: GENERAL INFORMATION:  
: APPLICANT: SASAKI, Keiko  
: APPLICANT: MORT, Takayuki  
: APPLICANT: MAKINO, Satoshi  
: TITLE OF INVENTION: ATTENUATED MEASLES VIRUS VACCINE.  
: TITLE OF INVENTION: CONTAINING SPECIFIC NUCLEOTIDE SEQUENCE AND A METHOD FOR  
: TITLE OF INVENTION: ITS ABSOLUTE IDENTIFICATION  
: NUMBER OF SEQUENCES: 19  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: YOUNG & THOMPSON  
: STREET: 745 South 23rd Street  
: CITY: Arlington  
: STATE: Virginia  
: COUNTRY: USA  
: ZIP: 22202  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: Patentin Release #1.0, Version #1.30  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/905,817  
: FILING DATE: 04-AUG-1997  
: CLASSIFICATION: 435  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 08/348,891  
: FILING DATE: 25-NOV-1994  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: US 07/848,400  
: FILING DATE: 10-MAR-1992  
: PRIOR APPLICATION DATA:  
: APPLICATION NUMBER: JP 3-293625  
: FILING DATE: 14-OCT-1991  
: ATTORNEY/AGENT INFORMATION:  
: NAME: PATCH, Andrew J.  
: REGISTRATION NUMBER: 32,925  
: REFERENCE/DOCKET NUMBER: KP-7501A  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: 703-521-2297  
: TELEFAX: 703-685-0573  
: TELEX: 248425 EMBON  
: INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 15894 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 108..1682  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1807..3327  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 3438..4442  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 5458..7107  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 7271..9121  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 9234..15782  
US-08-905-817-1  
alignment\_scores:  
Quality: 91.00 Length: 379  
Ratio: 0.587 Gaps: 18  
Percent Similarity: 40.897 Percent Identity: 20.580  
alignment\_block:  
US-09-438-185-1047 x US-08-905-817-1 ..  
Align seg 1/1 to: US-08-905-817-1 from: 1 to: 15894  
57 GlnAlaLeuAlaArgHisLysCysIleSerIleLeuGluPhePheLysAs 73  
10518 GAAGGTTAACACATGAGCAGTGGTTGATAACTGGAATCTTTGCTGG 10567  
73 nLeuLeuPheValHisLeuLeuSerLysAsnGlnArgGluGlyC 90  
10568 AGTGAATT.....GGCT 10581  
90 ysSerThrAspMetAlaValSerThrProPhePheAsnArgAsnLeu 106  
10582 GCTTTATGCTCTTAGCCTGGATAGTAT.....CTGACATGTACCTA 10625  
107 TtpTyrArgLeuLeuSerSerArgPheSerLeuTyrLysSerTyrCysPr 123  
10626 AAGGACAGGCGACTTGCCTCTCCAAAGGAATGGGATTCAGTTACCC 10675  
123 oArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheL 140  
10676 GAAAGAGTTCTCTGCTTACGACCTCCCAAGGAGACCGGTC..... 10718  
140 euAspHisGlnAlaValIleLysPheGluLeuGluThrHisPheSer 156  
10719 .....CGGAGGCTTGTAGATGTTTCTTAAATGATTCGAGCTTGC 10760  
157 TyTyrProValSerGlyPheValAlaProHisGlnTyrLeuSerLeu 173  
10761 CCATATCATGTGATAATGATGTTGTAAGTGGAGCTTAC.....CT 10801  
173 uGlnAspArgTyrPhePheProIleAlaSerValMetArgThrLeuAspLys 190  
10802 CCAAGACCTGACTCAACCTG..... 10823  
190 spAsnPheSerLeuThrProAspLeuIleHisAspLeuGlyHisVal 206  
10824 ..TCTTACGCTTGAAGAAAGAGATCAAGGAA..... 10856



119 sSerTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuL 136  
1952 ATTATAT .....CATAATTGCTTTATTGGCAAC 1980  
136 euSerAspPheLeuAspHisGlnAlaValIleLysPheGluLeuGlu 152  
1981 TAGACTCTATAGATCGTTGGTCAGCATGTAAGAGATATAAGGATCG 2030  
153 ThrHisPheSerTyrProVal ..... 160  
2031 CTTTCATTGAA .....CCCATGATACAATTTTCAAGAGCTATTAC 2074  
161 .....SerGlyPheValAlaProHis .....GlnTyrL 170  
2075 TTTTGACACAGAGAGTGTACTTACCAGTCGATTTCCCTCTTACA 2124  
170 euSerLeuGlnAspArgTyrPheProIleAlaSerValMetArgThr 186  
2125 AGTCAATATCGAAGATACTTACTAGTTGGGACGAGTCTCCCTCG 2174  
187 LeuAspLysAspAsnPhe...SerLeuThrProAspLeu.....I 199  
2175 CTTGATAAGAAATTTATGATCTCTTTCTGGAGATTTGGATAAAATAT 2224  
199 eHisAspLeuGlyHisValProTrpLeuLeuHisPro..... 212  
2225 GGTCTCGGTACTGGTCTGAGCTATTCAAGCTTTATCGTGAGCGAATATGA 2274  
213 .....SerPheSerGluPhePheIleAsnMetGlyArgLeuPheThrLys 227  
2275 CTTCAACATTTACGATTTCTACATTTGCTGAGAGAAACATTACCAAAA 2324  
228 .....ValIleGluLysValGlnAlaLeuProSerLysLysGlnArgI 242  
2325 GAGGAATATTAATTTCTAAGAAAGATCCCTCCACACCAACTCTT 2374  
242 eGlnThrLeuGlnSer.....AsnLeuIleAlaI 252  
2375 AGAATAGACAGAAACACCGAGCATTTGCAAGTAGCAGCACTAATTTAT 2424  
252 leValArgCysPheThrPheThrValLysSerGlyLeuIleGluAsnHis 268  
2425 TCATGCAACCAATCTTCCTCCCTTTGAAACCATGGTCTCATTAAG...TTT 2471  
269 GluGlyArgLysAlaTyrGlyAlaVal 277  
2472 CAAAGCACCAGAGTTAGCATCTGGTA 2498

seq\_name: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq:us-08-484-106-5

documentation\_block:

Sequence 5, Application US/08484106

Patent No. 5614618

GENERAL INFORMATION:

APPLICANT: STILLMAN, Bruce

APPLICANT: BELL, Stephen P

APPLICANT: KOBAYASHI, Ryuji

APPLICANT: RINE, Jasper

APPLICANT: FOSS, Margit

APPLICANT: McNALLY, Francis J

APPLICANT: LAURENSEN, Patricia

APPLICANT: HERSKOWITZ, Ira

APPLICANT: Li, Joachim J

APPLICANT: GAVIN, Kimberly

TITLE OF INVENTION: ORIGIN OF REPLICATION COMPLEX GENES

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESS: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT

STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: USA

ZIP: 94111-4187

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/08/484,106  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Osman Ph.D., Richard Aron  
REGISTRATION NUMBER: 36,627  
REFERENCE/DOCKET NUMBER: A-59032/DJB/RAO  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 494-8700  
TELEFAX: (415) 494-8771  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2700 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-484-106-5

alignment\_scores:  
Quality: 90.50 Length: 309  
Ratio: 0.584 Gaps: 16  
Percent Similarity: 50.162 Percent Identity: 20.712

alignment\_block:  
US-09-438-185-1047 x US-08-484-106-5 ..  
Align seg 1/1 to: US-08-484-106-5 from: 1 to: 2700

24 LeuSerLeuPhePheGlnAsnSerGlnSerLeuGlnArgAlaTyrSerTh 40  
1650 AATGCTACTTTTCCAGAAATGCTTTTCAGTA.....TTCAATGA 1690  
40 rProTyrSer.....TyrTyrArgIleIleLeuG 50  
1691 CCCTCTAATGCTGATTTTGAACGACGACTACTTAAATACTG... 1736  
50 LysGluAsnLysGluLysGlnAlaLeuAlaArgHisLysCysIleSer 66  
1737 .....AGCAGATGCTCTACA 1751  
67 IleLeuGluPhePheLysAsnLeuPheValHis..... 78  
1752 TTTATGTTCTTTCGAGAGGCTTTTAAACGACGATGCTCTGCTGACGA 1801  
79 LeuLeuSerLeuSerLysAsnGlnArgGluGlyCysSer.....T 92  
1802 AATTTCTTCATATTGCAACAAACAAACAGAGCCCTAGAGAGATTTTG 1851  
92 hrAspMetAlaValValSerThrPro.....PhePhe 102  
1852 TTGAGTTTGTGTAAGAGAGAACCCGATTAAACGGCGATGCTAAGTTTGT 1901  
103 AsnArgAsnLeuTyrTyrArgLeuLeuSerSerArgPheSerLeuTrpL 119  
1902 GCTCGATTCTCTCGAAGAAGAAATTAATATAACCAATTTTAACTCTGATAGA 1951  
119 sSerTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuL 136  
1952 ATTATAT .....CATAATTTGCTTATTGGCAAC 1980  
136 euSerAspPheLeuAspHisGlnAlaValIleLysPheGluLeuGlu 152  
1981 TAGACTCTATAGATCGTTGGTGGTCAAGATGTAAGAGATATAAGGATCGG 2030



```

31491 TCATGCAAGCAATCTTCGCTTTGAAACATGGGTCTCATTAAG...TTT 31537
      ::::: |||||||:::
269  GUGUlyArGlySAlyArGlyAlaVal 277
      ::::: |||||||:::
31538 CAAGCACCAGAGATTACGATCTGGTA 31564
      ::::: |||||||:::

seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq;US-09-434-774-9

seq_documentation_block:
  Sequence 9, Application US/09434774A
  Patent No. 6235514
  GENERAL INFORMATION:
  APPLICANT: Croteau, Rodney B
  TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING ISOPENTENYL
  TITLE OF INVENTION: MONOCHOSPHATE KINASE, AND METHODS OF USE
  FILE REFERENCE: wslr14448
  CURRENT APPLICATION NUMBER: US/09/434,774A
  CURRENT FILING DATE: 1999-11-04
  NUMBER OF SEQ ID NOS: 16
  SOFTWARE: PatentIn Ver. 2.0
  SEQ ID NO 9
  LENGTH: 1203
  TYPE: DNA
  ORGANISM: Lycopersicon esculentum
  FEATURE:
  NAME/KEY: CDS
  LOCATION: (1)..(1203)
  US-09-434-774-9

alignment_scores:
  Quality: 89.50 Length: 349
  Ratio: 0.963 Gaps: 19
  Percent Similarity: 45.559 Percent Identity: 22.923

alignment_block:
US-09-438-185-1047 x US-09-434-774-9 ..
Align seg 1/1 to: US-09-434-774-9 from: 1 to: 1203

68  LeuGluPhePheLysAsnLeuLeuPheValHisLeuLeuSerLeuSerly 84
    ||| ::::: |||||||::: |||
43  TTGATTCCTTTTAAAGAGATGTTGTTTTTCAGTCATGGTCA..... 84
    ||| ::::: |||||||::: |||
84  sAsnGlnArgGluGlyCysSerThrAspMetAlaValSerThrProp 101
    ||| ::::: |||||||::: |||
85  AATAGGCCACATGGTTCATCC.....T 106
    ||| ::::: |||||||::: |||
101  hePheAsnArgAsnLeuTyrTrpArgLeuLeuSer..... 112
    ||| ::::: |||||||::: |||
107  ATTTTAAACAAGAATATCAATTTTAGAAGAAACAGTTTGTATTGTGAAG 156
    ||| ::::: |||||||::: |||
113  .....SerArgPheSerLeuTyrLysSerTyrCysPr 123
    ||| ::::: |||||||::: |||
157  GCTTCAGGTTCAAGAACTAGTAAAAACAAGTAGAGATAACATATAATCC 206
    ||| ::::: |||||||::: |||
123  oArgPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPhe 140
    ||| ::::: |||||||::: |||
207  TGA.....GAGAAGTTTAAATAATGCTGATGAAG 238
    ||| ::::: |||||||::: |||
140  euAspHisGlnAlaValIleLysPhePheGluLeuCluThrHisPheSer 156
    ||| ::::: |||||||::: |||
239  TGGATAGAGAAGCTGGGCTTTCAAGACTCACTCCT.....TTTCT 279
    ||| ::::: |||||||::: |||
157  TyrTyrProValSerGlyPheValAlaProHisGlnTyrLeuSerLeu 173
    ||| ::::: |||||||::: |||
280  CTTTGCAGATAAATGTTTCTTG.....AGATAACAACGAAGAG 320
    ||| ::::: |||||||::: |||
173  uGlnAspArgTyrPheProIleAlaSerValMetArgThrLeu..... 187
    ||| ::::: |||||||::: |||
321  GGATGACGGATATCATGATTGGGCGTCTCTCTTCATGATTAATGACTGAC 370
    ||| ::::: |||||||::: |||

```

Mon Jul 2 08:46:43 2001

us-09-438-185-1047.rni

Page 18

...



OM of: US-09-438-185-1047 to: GenEmbl.\* out\_format : pfs  
 Date: Jun 29, 2001 2:47 PM  
 About: Results were produced by the GenCore software, version 4.5,  
 Copyright (c) 1993-2000 CompuGen Ltd.

## Command line parameters:

-MODE=frame-p2n.model -DEV=xlp  
 -Q/cgn2.1/USPTO\_spool/US09438185/runat\_29062001\_074544\_6960/app\_query.fasta\_1.430  
 -DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000  
 -MINMATCH=0.100 -LOOPCL=0.000 -LOAPEXT=0.000 -QGAPOP=4.500  
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
 -DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi  
 -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0  
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
 -MAXLEN=200000000 -USER=US09438185 -CGN1\_1\_5400 -NCPU=6  
 -ICPU=3 -LONGLOG -NO\_XLPXY -WAIT -THREADS=1

## Search information block:

Query: US-09-438-185-1047  
 Query length: 362  
 Database: GenEmbl.\*  
 Database sequences: 1344157  
 Database length: -85606004  
 Search time (sec): 1336.620000

## score\_list:

Sequence	Strd Orig	zscore	EScore	Len	Documentation
gb_ba1:AE001685	+ 1889.00	2691.87	1.1e-141	10847	AE001685 Chlamydia pneumoniae
gb_ba1:AE002240	+ 1889.00	2684.54	2.7e-141	10847	AE002240 Chlamydia pneumoniae
gb_ba1:AP002548	+ 1889.00	2657.85	8.3e-140	325865	AP002548 Chlamydia pneumoniae
gb_ba1:AE004410	+ 256.50	341.08	9.2e-11	10143	AE004410 Vibrio cholerae
gb_ov:GGA251387	+ 237.50	331.97	3.0e-10	1633	AJ251387 Gallus gallus mRNA
gb_ov:RRT2	+ 236.50	321.88	1.1e-09	3877	X53501 Rat mRNA for tryptophan
gb_ba1:AE004522	+ 236.50	307.88	6.5e-09	15724	AE004522 Pseudomonas aeruginosa
gb_ba1:HSWYDR	+ 231.50	325.34	9.4e-10	1335	X52836 Human mRNA for tryptophan
gb_ov:GGG26428	+ 231.50	322.96	9.4e-10	1757	U26428 Gallus gallus tryptophan
gb_ov:MUSPRA	+ 230.50	321.29	1.2e-09	1732	J04758 Mouse tryptophan
gb_ov:PEPPI	+ 230.50	316.34	2.2e-09	2843	M88627 Pseudomonas aeruginosa
gb_ov:RATTOHAB	+ 229.50	322.46	1.0e-09	1335	L29305 Homo sapiens tryptophan
gb_ov:RATTOHAB	+ 229.50	321.90	1.1e-09	1412	L29305 Homo sapiens tryptophan
gb_ov:RATTOHAB	+ 229.50	321.90	1.1e-09	2077	M2478 Quail tryptophan
gb_ov:RATTOHAB	+ 229.50	318.04	1.8e-09	1757	M63200 Mouse tryptophan
gb_ov:RATTOHAB	+ 229.50	313.95	3.0e-09	1757	M63200 Mouse tryptophan
gb_ov:RATTOHAB	+ 229.50	310.73	4.5e-09	2425	AJ000731 Anguilla anguilla mRNA
gb_ov:RATTOHAB	+ 229.50	312.01	3.8e-09	1847	AF031034 Schistosoma mansoni
gb_ov:RATTOHAB	+ 229.50	311.69	4.0e-09	1651	L20679 Xenopus laevis tryptophan
gb_ov:RATTOHAB	+ 229.50	310.27	4.8e-09	1770	M10244 Rat tryptophan
gb_ov:RATTOHAB	+ 229.50	308.51	6.0e-09	2111	L22651 Rat tryptophan
gb_ov:RATTOHAB	+ 229.50	308.54	6.0e-09	1696	M36794 Bovine tryptophan
gb_ov:RATTOHAB	+ 221.50	304.90	9.6e-09	2441	M17250 Rabbit tryptophan
gb_ov:RATTOHAB	+ 221.50	304.90	9.6e-09	332635	AP03005 Mesorhizobium loti
gb_ov:RATTOHAB	+ 220.00	305.70	8.1e-09	1816	X05230 Human mRNA for tyrosine
gb_ov:RATTOHAB	+ 220.00	305.29	9.1e-09	1891	Y00414 Human mRNA for tyrosine
gb_ov:RATTOHAB	+ 220.00	305.14	9.3e-09	1921	M17589 Human tyrosine
gb_ov:RATTOHAB	+ 219.00	305.21	9.2e-09	1650	X76209 D.melanogaster mRNA for
gb_ov:RATTOHAB	+ 216.50	302.61	1.3e-08	1494	AF030336 Schistosoma mansoni
gb_ov:RATTOHAB	+ 214.00	301.51	1.5e-08	1163	K02599 Rat phenylalanine
gb_ov:RATTOHAB	+ 214.00	293.43	3.0e-08	2429	K03020 Human phenylalanine
gb_ov:RATTOHAB	+ 213.50	292.45	4.2e-08	2679	L47726 Homo sapiens phenylalanine
gb_ov:RATTOHAB	+ 213.50	292.44	4.7e-08	2680	U49877 Homo sapiens phenylalanine
gb_ov:RATTOHAB	+ 212.50	294.14	3.8e-08	1706	X51942 Mouse mRNA for phenylalanine
gb_ov:RATTOHAB	+ 211.00	272.19	6.3e-07	12277	M36705 Bovine tyrosine
gb_ov:RATTOHAB	+ 210.00	294.08	3.8e-08	1706	X51942 Mouse mRNA for phenylalanine
gb_ov:RATTOHAB	+ 207.00	288.69	7.6e-08	1529	M32802 D.melanogaster phenylalanine
gb_ov:RATTOHAB	+ 206.50	287.26	9.2e-08	1642	AF135186 Caenorhabditis elegans
gb_ov:RATTOHAB	+ 197.00	274.82	4.5e-07	1449	M81833 Drosophila melanogaster
gb_ov:RATTOHAB	+ 196.50	278.93	2.7e-07	894	AF146711 Chromobacterium violaceum
gb_ov:RATTOHAB	+ 193.00	235.80	6.8e-05	40392	U21308 Caenorhabditis elegans

gb\_ba1:AE005836 - 189.50 242.72 2.3e-05 12203 AE005836 Caulobacter cresc  
 gb\_ba2:CHRPHEHYD - 187.00 262.25 2.3e-06 1207 M55915 Chromobacterium viol  
 gb\_in3:DMPEHY - 182.00 243.62 2.5e-05 3786 X98116 D.melanogaster pah/t  
 gb\_in3:AC019368 - 182.00 210.64 0.0017 102552 AC019368 Drosophila melan

seq\_name: gb\_ba1:AE001685

## seq\_documentation\_block:

LOCUS AE001685 10847 bp DNA BCT 01-DEC-2000  
 DEFINITION Chlamydia pneumoniae section 101 of 103 of the complete genome.  
 ACCESSION AE001685 AE001363  
 VERSION AE001685.1 GI:4377378  
 KEYWORDS  
 SOURCE Chlamydia pneumoniae CWL029.  
 ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.  
 REFERENCE 1 (bases 1 to 10847)  
 AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Hyman,R.W.,  
 Olinger,L., Grimwood,J., Davis,R.W. and Stephens,R.S.  
 TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis  
 JOURNAL Nat. Genet. 21 (4), 385-389 (1999)  
 MEDLINE 99206606  
 PUBMED 10192388  
 REFERENCE 2 (bases 1 to 10847)  
 AUTHORS Kalman,S., Mitchell,W., Marathe,R., Lammel,C., Fan,J., Olinger,L.,  
 Grimwood,J., Davis,R.W. and Stephens,R.S.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University  
 of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA  
 FEATURES  
 source  
 1..10847  
 /organism="Chlamydia pneumoniae CWL029"  
 /strain="CWL029"  
 /db\_xref="taxon:115713"  
 complement(134)..826  
 /gene="Cpnl045"  
 complement(134)..826  
 /gene="Cpnl045"  
 /codon\_start=1  
 /transl\_table=1  
 /product="conserved hypothetical bacterial membrane protein"  
 /protein\_id="AADI9182.1"  
 /db\_xref="GI:4377378"  
 /translation="MNTSHKTLVFSYLSSTFLLVLSNLSKLTPTTFNFIIP  
 GGLIYPTLISDVVNEIFGPKARVIFSAFIAANLASSIVQIFMFPVSPMOT  
 ANHCLFDLSPRLASLLAIVSQDLIVLYTFKRNTPNSLWLRNSGTSWISQIPD  
 TIVTQCILYFGMGLSPQTLNIMFYIYKTFVCLVTLPLFLAVNTIRKFLGMPST  
 KIANVPLINP"  
 complement(823)..1911  
 /gene="Cpnl046"  
 complement(823)..1911  
 /gene="Cpnl046"  
 /codon\_start=1  
 /transl\_table=1  
 /product="Aromatic Amino Acid Hydroxylase"  
 /protein\_id="AADI9183.1"  
 /db\_xref="GI:4377380"  
 /translation="MHYCETDLPKYLKALKQSLFQNSQSLQRAYSTPYSY  
 YRIILQNKQKQALAKKICILEFFKNLLFVHLLSKNRCQECSTDMAYVSTPFF  
 NRIYRLKNSRFSKWSYCPREFLDYLFGLSDFLDQAVIKFPELETHSYTFP  
 SGVAFPHOYLSLDQRYFPFASVRLDKDNFLDHLHGLVPLHPSSEFF  
 INMGRLFKVLEKQALPQKQRIQIQNLAIKRCFWFTVSGELNHEGRKAYG  
 VLISSPQELGFLPNDVRLVLELDLIIRPFTPTQPTLFRHDFELVELTSKLE  
 WMLDQCLLESIPLYNQEKYLSGFEVLCQ"  
 1788..2582  
 /gene="daps"  
 /note="Cpnl047"  
 1788..2582  
 /gene="daps"  
 /codon\_start=1  
 /transl\_table=1  
 /product="Dihydrodipicolinate Reductase"

```

/protein_id="AAD19184.1"
/db_xref="GI:4377381"
/translation="MGSSMHVGVIGCSGRTGKIVSALEOSSEYTLGPGFSRSSALTL
FOVIAHNDLVDFSHPLTKEVVVALLISPKPLIIGTTPGPKCKEAHSDLEELTHIV
PVVVCNAGSLGAYIKRLVLMQLSQCNPQFDIRIRETHRYKDKSLGSAQDLDTIQ
QVKQEDWGEYEVGQDSSKTIIEVQSSRVGDIPIGEHVAFISGGQILVRHTVFSRN
VFGRLSLILDKLNTLNPQGLYSLGDTLELVLRNEHCLLKKTTDH"
/gene="asd"
/notes="Cpnl048"
/translation="MRIAVALGVTLGVQGFVALLHKWYRDVIAEVAASNKYGQSYG
DACIMQEPGPMPEWRDLPINKIEVQSDIVVSFLPSSAESMEAYCISQGVVFSNA
STYRMSSVPIIIPVNSDHFOLLEQPIYKGIITSPCCVSGITLALAPLRKFLSDH
VHIVLOSAGAGYPGVPSLDLANTVPHIVGEEKILRETVKILGSDNQPLPKLSV
TVHRVPVAYGHTLSLVTFKDVLDLDELILSYQEKNEFPNTVQILYDNPWPQARKHL
SHDDMRVHLGPTIGGDFRTKMNVLIHNLVRGAAGTLLASMENYFPDYLRKEMCLR"
/3831..5153
/gene="lysC"
/notes="Cpnl049"
/3831..5153
/gene="lysC"
/codon_start=1
/transl_table=11
/product="Aspartokinase III"
/protein_id="AAD19186.1"
/db_xref="GI:4377383"
/translation="MSKIVYFGTSLATFNICLVCDIICKDKSFVVVSAIAGVTD
LYVDCSSLSIREEVLEKIEGHEEVKNAIPFPVSWTWSRLPYLQHLIEISLDF
ARLSGEDISLVRACVSGWDLGFLEARSVILIDUSYRASPNLDMKAHHQOL
ELMQPSTIQGIGSGEVLGIVGEGSDISATILAEARATFVIRTVDNVGIITMD
YVSDQRIPELSPFEMONLAKGAKVLPPLFCMRAGIPFIVFTFDEKRGTW
PIAVDSVSEPRIKALSQSQSCVDVTVLGCGLGEEILGIESHIDFELMIAQ
NNVGVPMDDCIISQEAQEHLDVLSLSSVTLRLHSVALITMIGDNLNLSKPFVSVTIE
5162..6037
/gene="dapA"
/notes="Cpnl050"
5162..6037
/gene="dapA"
/codon_start=1
/transl_table=11
/product="Dihydrodipicolinate Synthase"
/protein_id="AAD19187.1"
/db_xref="GI:4377384"
/translation="MHLITATVTFPPNGTIDFASLERLLSFODAVNGVVLGSGTE
GLSLTKKEQALICFACDLQKLPFLVGTSTLLEVLDMWIFHCNDLPISGFLMTPI
YTPKLCGQILNFEAVLNAAKHPALYNIPSRATPLYLDTVKALAHPOFTLGKDSG
GSVEEFQSKYSTAPHIOLYCGDDVFWSEMAACGAGHLISLVSNAPFEAREVLPQE
QDYSLSLMTCTRWVTTNPIGIKAILAYKKAITHAQLRLPLSIEDFDLENVPVAVES
MLANPKLRVSFVSYS"
/6198..6509
/gene="Cpnl051"
/6198..6509
/gene="Cpnl051"
/codon_start=1
/transl_table=11
/product="hypotheical protein"
/protein_id="AAD19188.1"
/db_xref="GI:4377385"
/translation="MTPKSIQQLHLIKTIDFVRKISFVTTKKSSFFRQSLRFLFLFW
MFLYICRSIRFCHVIATFICRGILFLITLFLSMICILHIFITLFWICKEDPRIIRKN
K"
/6553..7408
/gene="Cpnl052"
/6553..7408
/gene="Cpnl052"

```

```

/codon_start=1
/transl_table=11
/product="hypotheical protein"
/protein_id="AAD19189.1"
/db_xref="GI:4377386"
/translation="MKYNSREKIKSALRITCSYCIYVFRNNSLSYCDKIFYSLSCYV
FNGNSLIGRCRSCFCFRGKKTETVEKVKIDKIRPLEGNDPVKVAESFPKRRAALE
SLSSQSIGNLCAISNFDLSQMLSRNFSKEIWSGTITFRSKSTCDAEGSPFRYTAGC
YLAGRLSKLAGSEYELGVTAGLQGLRQKDVSDSHRTRATSSILSVHGSMTVRPLSCTKY
IVGKARPLLFPRLLTSDVRRDLKKRFLRECKD"
/7389..7940
/gene="Cpnl053"
/7389..7940
/gene="Cpnl053"
/codon_start=1
/transl_table=11
/product="hypotheical protein"
/protein_id="AAD19190.1"
/db_xref="GI:4377387"
/translation="MNFADIDHNLHLYLCLGLDGVACPILLSLTDCLPNYSKASHEVLV
YSKFRCSGSEPSRLATSGNDTYISIVLPIGLRYEVTSPSGRHDNDIMHVAPKIGV
LSHGTREAKEIPGSSKDYAFFSLTARESLMISEKLAMTFOVSEVQNCYSQCTKVTKT
NLKEQYRHLSHNTGFELS VKSAP"
/8270..10705
/gene="Cpnl054"
/8270..10705
/gene="Cpnl054"
/codon_start=1
/transl_table=11
/product="hypotheical protein"
/protein_id="AAD19191.1"
/db_xref="GI:4377388"
/translation="MOVLLSPOLPPPHSVGSISSPSKLRVLAINFLVFGMLLISL
ALFLTICPGLSAAISVGLICIGLSALGVLMISGLICLLKVRLEIPVRPEIPEGVSL
APSEPALQAOKTALQPKELDQDLTIOEVFACILKDKSKESRSLNDAKKELR
VDFVWEDILSELIFELQIYVQEGWDLNGLNGSLAMTASESDLDLFFVSKRLGYL
PSGDVIGLSSAKKIEIVARLMSLHCEIHKVAVAFDRNSTAMAEKAFKALGALRESV
YRSLTYSYRDLFESERAKIPWNGHITWLDODAKSGCAERKADKRDARWAKFKRAVFW
VEEDGGDINLLGDSGTVIDPYRQRMDEITFHELYEKTFFLKRHRKCALAKTTFE
KKRSNKLAYAEIENYRCVQREDEYWKKEERAEFRERGNKLLSPSELESLEQFD
TRSNLEKAYAEIENYRCVQREDEYWKKEERAEFRERGNKLLSPSELESLEQFD
HGLNFSEKLEMELEHILKQAEATAVENKILSDASERLEIVFEDVKMERVSLRDL
KTLRMAELPLLPKKAPEKACSOYNCAEMLEKVKPYCKESLAYVTSKERVLSLDEL
RRAYTECQRFQDGSLESEVACREQLREIQEFETQGLDLVEKLLCVCYSLRNTDE
CDCVSGYKKEAPECKKFAQYDEIYRVVQSRWMTSERLRGEGVQACNMKMLKALUSE
BDKVLKEEYWLRYERKNKRLVGTQIVATQORVAAFESIEVPEIPEAPEKEPSLL
DKARSLFTREDHT"
/3261 a 1883 c 2436 g 3267 t
BASE COUNT
ORIGIN
alignment_scores:
Quality: 1889.00 Length: 362
Ratio: 5.218 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-438-185-1047 x AE001685/rev ..
Align seg 1/1 to reverse of: AE001685 from: 1 to: 10847
1 ValHistTyrCysGluArgThrLeuAspProLysTyrIleLeuLysIleal 17
1911 GTGCACCTACTGCGAGAGAACCTGACCCAAAGTATATCTGAAGATTGC 1862
17 aLeuLysLeuArgGlnSerLeuSerLeuPheGlnAAsnSerGlnSerL 34
1861 TCTAAAGCTGAGACAATCACTTTCCTCTCTCCAGAACAGCAATCAC 1812
34 euGlnArgAlaTyrSerThrProTyrSerTyrArgIleIleLeuGln 50
1811 TCCAAACGTGCATCTCGACCCCATATTCTCTACTACCGAATCATCTACAA 1762

```

```

51  LysGluAsnLysGluLysGlnAlaLeuAlaArgHisLysCysIleSerI1 67
1761 AAGGAAATAAAGAGAGCAAGCTTTAGCTCGACAAATGCAATTTCTAT 1712
67  eLeuGluPhePheLysAsnLeuLeuPheValHisLeuLeuSerLeuSerL 84
1711 TTTAGAAATTTTCAAAACTTACTCTTTGTTTCATCTTCTGTCATTATCAA 1662
84  yAsnGlnArgGluGlyCysSerThrAspMetAlaValValSerThrPro 100
1661 AGAATCAAGAGGAGGTGCTGCACATGATATGGCTGTGTGAAGCACTCC 1612
101 PhePheAsnArgAsnLeuTrpTyrArgLeuLeuSerSerArgPheSerLe 117
1611 TTTTAAATCGGAATTTATGGTATCGACTCTCTTTCTCACGGTTTCTCT 1562
117 uTrpLysSerTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheG 134
1561 ATGGAAGCTATGTCCAAAGATTTTCTTGATTTACTTAGAAGCTTTCG 1512
134 yLeuLeuSerAspPheLeuAspHisGlnAlaValIleLysPhePheGlu 150
1511 GTCTCTTCTTGATTTTGTAGACCAACAGCACTTAATTAATTTTCGAA 1462
151  LeuGluThrHisPheSerTyrTyrProValSerGlyPheValAlaProH 167
1461 TTAGAAACACATTTTCTTATATCCGTTTCAGGATTTGTAGCTCCCA 1412
167 sGlnTyrLeuSerLeuLeuGlnAspArgTyrPheProIleAlaSerVal 184
1411 TCAATACATGTCTGTGTGAGGACCGTTTACATTTCCCATTTGCCTCTGTAA 1362
184 eArgThrLeuAspLysAspAspPheSerLeuThrProAspLeuIleHis 200
1361 TCGCAACTCTCGATAAAGATAATTTCTCTTAACCTCGATCTCATCCAT 1312
201 AspLeuLeuGlyHisValProTyrLeuLeuHisProSerPheSerGluPh 217
1311 GACCTTTTAGGACGTCGCTTGGCTTCTACATCCCTCATTTTCTGAAT 1262
217 ePheIleAsnMetGlyArgLeuPheThrLysValIleGluLysValGlnA 234
1261 TTTCAATACATGGAGAGACTTCTACTAAATCATATAAAGCAATCTGATC 1212
234 laLeuProSerLysLysGlnArgIleGlnThrLeuGlnSerAsnLeuIle 250
1211 CTCTTCTTAGTAAACCAACGATCAACCCCTACAAGCAATCTGATC 1162
251 AlaIleValArgCysPheThrPheThrValGluSerGlyLeuIleGluAs 267
1161 GCTATTGTACGCTGCTTTGTTGTTACTGTGTAAGCGGACTTATGAAAA 1112
267 nHisGluGlyArgLysAlaTyrGlyAlaValLeuIleSerSerProGlnG 284
1111 CCATGAAGCAAGAAAGCATATGAGCCGCTTCTTATCAGTTCTCTCCAG 1062
284 luleGlyHisAlaPheIleAspAsnValArgValIleProLeuGluLeu 300
1061 AACTTGACACGCTTTCATTGATAGTACGTCGTTCTCCCTTTAGAATTG 1012
301 AspGlnIleIleArgLeuProPheAsnThrSerThrProGlnGluThrLe 317
1011 GATCAGATTTATTCGCTTCCTTCAATATACATCACTCCACAGAGACTTT 962
317 uPheSerIleArgHisPheAspGluLeuValGluLeuThrSerLysLeuG 334
961 ATTTCATAGACATTTTGAAGACGTTGAGAACCTCACTTCAAAATTAG 912
334 luTrpMetLeuAspGlnGlyLeuLeuGluSerIleProLeuThrAsnGln 350
911 AATGGATGCTGACCAAGGCTGTGTAGAAATCAATCCCTTTTACAAATCAA 862
351 GluLysTyrLeuSerGlyPheGluValLeuCysGln 362

```

861 GAGAAATATCTTCTGTTGTTGAGTACTTTTGCAA 826

seq\_name: gb\_bai:AE002240

seq\_documentation\_block: 22589 bp DNA BCT 30-MAY-2000  
 LOCUS AE002240 Chlamydomophila pneumoniae AR39, section 68 of 94 of the complete  
 DEFINITION genome.

ACCESSION AE002240 AE002161  
 VERSION AE002240.2 GI:8163502

KEYWORDS

SOURCE

ORGANISM

Chlamydomophila pneumoniae AR39.

Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.

REFERENCE 1 (bases 1 to 22589)

AUTHORS Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Winn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.

Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39

Nucleic Acids Res. 28 (6), 1397-1406 (2000)

JOURNAL

MEDLINE

PubMed

REFERENCE

AUTHORS

2 (bases 1 to 22589)

Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Winn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.

Direct Submission

Submitted (01-MAR-2000) The Institute for Genomic Research, 9712

Medical Center Dr, Rockville, MD 20850, USA

On Jun 1, 2000 this sequence version replaced gi:7189730

COMMENT

gi:7189720.

FEATURES

Source

1..22589

/organism="Chlamydomophila pneumoniae AR39"

/strain="AR39"

/db\_xref="taxon:115711"

/note="synonym: Chlamydia pneumoniae AR39"

complement(161..955)

/gene="CP0805"

complement(161..955)

/gene="CP0805"

/note="similar to GB:D10483 SP:P04036 GB:M10611 PID:145710

PID:216458. Identified by sequence similarity; putative"

/codon\_start=1

/product="dihydrodipicolinate reductase"

/protein\_id="AAF38603.1"

/db\_xref="GI:7189721"

/translation="MGSSMHVGVIGSGRTGKVIYSALEOSSEVTLGGFSSALTL

FVIAHNDVLVDFSHPLITREVAHLHLISFKELIIGTTFPKCKEADSLBELTHIV

PVVCNPSALGAYTHKRLVMLLSQCLNCPQDIRIRETHRRKADSLGTAQLDLDTIQ

QVKQEDGEEYEGVORDSRKKTIEVSSRVGDIPEGHEVAFISSEQLVHRHVFNRN

VFGRIILSLDLWLKTLNPQPGLYSLGDTLELRLRHEHCLLAKTKTDH"

832..1920

/gene="CP0806"

832..1920

/gene="CP0806"

/note="biopterin-dependent aromatic amino acid hydroxylase

family protein; identified by match to PFAM protein family

HMM PF00351"

/codon\_start=1

/transl\_table=11

/product="biopterin-dependent aromatic amino acid

hydroxylase family protein"

/protein\_id="RAF3705.1"

/db\_xref="GI:8163503"

/translation="MHYCETLDPKYILKILKRLQSLSPFQNSQSLQRAYSTPTYSY



```
832 GTGCACACTACTGCGAGAGAACCCCTGGACCCCAAGATATATATCTGAAGATTGC 881
17 aLeuLysLeuArgGlnSerLeuSerLeuPhePheGlnAsnSerGlnSerL 34
|||||
882 TCTAAAGCTGAGCAATCACTTCTCCCTGCTTCTCCAGACAGCAATAC 931
|||||
34 euGlnArgAlaTyrSerThrProTyrSerTyrArgGlnLeuGln 50
|||||
932 TCCAAACGTGCATCTGACGCCCATATCTCTACTACCAATCATCTTACAA 981
|||||
51 LysGlnAsnLysGlnLysGlnAlaLeuAlaArgHisLysCysIleSerI 67
|||||
982 AAGCAAAATAAGAGAGAACGCTTTAGCTGCACACAAATGCATTTCTAT 1031
|||||
67 eLeuGluPhePheLysAsnLeuLeuPheValHisLeuLeuSerLeuSerL 84
|||||
1032 TTTAGATTTTCAAAACTTACTCTTTGTTTCATCTTCTGTCATTATCAA 1081
|||||
84 yAsnGlnArgGluGlyCysSerThrAspMetAlaValSerThrPro 100
|||||
1082 ACAATCAAGAGGAGGTGCTCCACTGATATGGCTGTGTGTAACACTCCC 1131
|||||
101 PhePheAsnArgAsnLeuTrpTyrArgLeuLeuSerSerArgPheSerLe 117
|||||
1132 TTTTAAATCGGAATATATGGATCGATCTCTTCTCAGCGTTTCTCT 1181
|||||
117 uTrpLysSerTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheG 134
|||||
1182 ATGGAAGAGCTATTTGTCGAAGATTTTCTTGATTTACTTAGAAGCTTTCG 1231
|||||
134 lYLeuLeuSerAspPheLeuAspHisGlnAlaValIleLysPhePheGlu 150
|||||
1232 GTCTCTTCTGATTTCTTAGACCAATCAAGCAGTCATTAAATCTTCGAA 1281
|||||
151 LeuGluThrHisPheSerTyrTyrProValSerGlyPheValAlaProHi 167
|||||
1282 TTAGAACAACATTTTCTTATTATCCGTTCTCAGGATTTGTAGTCCCCA 1331
|||||
167 sGlnTyrLeuSerLeuLeuGlnAspArgTyrPheProIleAlaSerValM 184
|||||
1332 TCAATACTTCTCTGTTGTCAGGACCGTTACTTCCCATTCGCTCTGTAA 1381
|||||
184 eArgThrLeuAspLysAspAsnPheSerLeuThrProAspLeuIleHis 200
|||||
1382 TCGGAACCTCTCGATAAGATAATTTCTCTTAACTTCCCTGATCTCATCAT 1431
|||||
201 AspLeuLeuGlyHisValProTyrLeuLeuHisProSerPheSerGluPh 217
|||||
1432 GACCTTTTAGGCGACGTGCTTGGCTCTTACATCCTCATTTCTGAAT 1481
|||||
217 ePheIleAsnMetGlyArgLeuPheThrLysValIleGluLysValGlnA 234
|||||
1482 TTTTCAAAACATGGGAAGACTTCTCAATAGCTCATAGAAAAGTACAAG 1531
|||||
234 lAlaLeuProSerLysGlnArgIleGlnThrLeuGlnSerAsnLeuIle 250
|||||
1532 CTCTCTCTAGTAAACAAACAGCATACAAACCTTACAAAGCAATCTGATC 1581
|||||
251 AlaIleValArgCysPheThrPheThrValGluSerGlyLeuIleGluAs 267
|||||
1582 GCTATTGTACGCTCTTTTGTGTTTACTGTGTAAGCGGCAATTTATTA 1631
|||||
267 nHisGluGlyArgLysAlaTyrGlyAlaValLeuIleSerSerProGlnG 284
|||||
1632 CCAATGAGGAAGAAACATATGAGCGCGTCTTATCAGTCTCTCTCAGG 1681
|||||
284 lLeuGlyHisAlaPheIleAspAsnValArgValLeuProLeuGluLeu 300
|||||
1682 AACTTGACACCTTCTTATGATACCTAGCTGTGTTCTCCCTTTAGATTC 1731
|||||
301 AspGlnIleIleArgLeuProPheAsnThrSerThrProGlnGluThrLe 317
|||||
1732 GATCAGATTATTCGTTCTCCCTCAATACATCAACTCCACAGAGACTTT 1781
```

```
317 uPheSerIleArgHisPheAspGluLeuValGluLeuThrSerLysLeuG 334
|||||
1782 ATTTTCAATAAGACATTTTGTGTAACCTGGTAGAATCACTTCAAAATTAG 1831
|||||
334 luTrpMetLeuAspGlnGlyLeuLeuGluSerIleProLeuTyrAsnGln 350
|||||
1832 AATGGATGCTCAGCAAGCTGTGTAGAACTATTCCTCTTACATCAA 1881
|||||
351 GluLysTyrLeuSerGlyPheGluValLeuCysGln 362
|||||
1882 GAGAAATATCTTCTGCTGTTGAGGTACTTTGCCAA 1917
seq_name: gb_ba2:AP002548

seq_documentation_block:
LOCUS AP002548 325865 bp DNA BCT 08-DEC-2000
Chlamydomydia pneumoniae J138 genomic DNA, complete sequence,
DEFINITION section 4/4.
ACCESSION AP002548 AB033788 AB033790 AB033791 AB033818 AB036083
AB036084 AB036086 AB036087 AB036088 AB036089 AB036350
AB038351 AB038352 BA000008
AP002548.2 GI:10176694
VERSION
KEYWORDS
SOURCE Chlamydomydia pneumoniae J138 (strain:J138) DNA.
ORGANISM Chlamydomydia pneumoniae J138
REFERENCE Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomydia.
AUTHORS Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,
Shiba,T., Ishii,K., Hattori,K., Kuhara,S. and Nakazawa,T.
TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA
JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)
REFERENCE
MEDLINE 20330349
2 (bases 1 to 325865)
AUTHORS Shirai,M.
TITLE Direct Submission
JOURNAL Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.
Mutsunori Shirai, Yamaguchi University School of Medicine,
Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi
755-8505, Japan (E-mail:mshirai@po.cc.yamaguchi-u.ac.jp,
Tel:81-836-22-2227, Fax:81-836-22-2415)
COMMENT On Sep 15, 2000 this sequence version replaced gi:6172302
gi:6172304 gi:6172306 gi:6172308 gi:6172400 gi:6635182 gi:6635184
gi:6635186 gi:6635188 gi:6635190 gi:6635192 gi:6635194 gi:8547442
gi:8547447 gi:8547451 gi:8979170.
AB033788-AB033791, AB033818: Submitted (25-Oct-1999)
AB038350-AB038352: Submitted (14-Feb-2000)
AB036083-AB036089: Submitted (18-Dec-2000).
FEATURES
source
1..325865
/organism="Chlamydomydia pneumoniae J138"
/strain="J138"
/db_xref="taxon:138677"
/note="Synonym:Chlamydia pneumoniae (strain J138)"
221..1318
/gene="CPj0797"
221..1318
/gene="CPj0797"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/db_xref="GI:8979171"
/translacion="MSKKIKVLGHLTCLTFLRGVLCALALNSIGYASTSQESPYQKSI
EDWKYTFDLELLSKESGEAHISGNSRIVGASGOGSVATVWESHILKHLGT
LGGEASSAGISKDEGVVGVSTREGYTHAFVDFGRMDKDLGTIGATYSVARGVSGD
GSIIIVGSATARGEDYQGVKWKGIKQLKLLPQLGSEANAISEDYTVIVGRGE
ISRNHIVAKWNKNVATSLGTGGSVASAEASANGKVIWGSTNNGETTHAFMHKDE
TWHDLGTGGFSVATGVSADGRAIVGFSVAVTGTBIHAFYAEEMEDTLTLCGEAR
VFDISSEGNLIIGSIKTDAGAENAYLFTHK"
1467..2480
/gene="CPj0798"
gene
CDS
```

CDS	1467..2480 /gene="cpj0798" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAA99006.1" /db_xref="GI:8979172" /translation="MKKCCNYRSGVWFSVLVFLTQTLFAGHFDIGTSGLYSW ARGSDGVVGVYEGGNAPKVDGEKFLLEGLVPSEALVFKASVDGSGVIGISDOD PSCRKVVNGALVDLGIFFSEGMSPFAGVSDDGKTLVGLCLASDDPRTNFAVKWDETC MYVLPNLPDRSCARDSEDSVIVGDAMGSEELAKANYKDKGQHLILSNLPCAKRS SAHVSXKDSGFIYGEISEENENAFYHNGYIKDITGLGGDYVATGVSRCKYVIG HSFTDCEHAKYVDGRMIDLGIGSASFATGVSDDGKTIIVGKFEIELGEGCHAFII LDD complement(2558)..3607 /gene="cpj0799" complement(2558)..3607 /gene="cpj0799" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="BAA99007.1" /db_xref="GI:8979173" /translation="MAAIKQILRSMLOSLLMVMVLFSLXSGYCYVITDKPEDDHS SSAVKWDHMKTTLSRLNKRKASAKAVSGTCAITGVGFIKDTWSRTYAVRWNYWGKEL PTSGWVKKATGISDSGLIAGIVENELQSFAVTKWNNMYLLPSTWAVOSKAYGI SDGSVIVGSANDRTFAVKWTGHEAQVLPVGNWAKSVANSVSANGSIIVGSVODA SGILYAVKNEGNTIHTLGTIGYSALAKAVSNCKVIVGRSETTYIGEYHAFCHKNVGM SDLGTGGSYSAKGVSGATGVIVGMSTANGKLHAPKYVGGRMIDLGYSWKACAN AVSIDGEIIVGVQSE" complement(3867)..5153 /gene="eno" complement(3867)..5153 /gene="eno" /codon_start=1 /transl_table=11 /product="enolase" /protein_id="BAA99008.1" /db_xref="GI:8979174" /translation="MFEAVIADIQAREILDSRGYPTLHVKVTITSTGSGVEARVPSGAS TGKKEALEFRDPSRYQCKGVQVAKNVKEILFPLVKGCSVYEQSLIDSLMDSGDS PNKRTAGANALIGSLATAHAAATLRPLRYVILGGCFACSLPCMNNLINGSMAHND GLGFQEMTRPIGASSIKREAVNMGADVFHTKLKLLHERGLSTGVGDEGFAFNLASNE EALLELLATEKAGTFCGKIDSLADCAASSFVNKVTGYDGRHYEEQIAILSNICDR YPIDSTEDGLADEYDGNALLTEVLGKQVIVGDLLFTVNPTELILEGISNGLANSVLI KPNQIGTLTETVATKLAQMAGYTTIISHRSGETTDTTADLAVAFNAQIKTGSLSR SERVAKYNLMIEEELGSEALFTDSNVFSEVDEE" complement(5345)..7318 /gene="uvrB" complement(5345)..7318 /gene="uvrB" /codon_start=1 /transl_table=11 /product="exonuclease ABC subunit B" /protein_id="BAA99009.1" /db_xref="GI:8979175" /translation="MTFQLHAPFACQDQPEATARLSAGVRNOVKSQVLLGTGSGKAT FTIANVAVNPLTLVAHNHTLAAQYQEFREFFPNNAVEFYIYDYQPEAYIAR SDYIEKSLILNDEIKLRLSATRSILERRDTLIVSVCYIGSPENYTSMAVLE VGEYEPNLTALQVLMHYQASPIQSFARFERSVSDIFPAYESEALARLEFLNDTL TSEYSDPMTIPKESVPSATLYPGSHYVIEPAIRQOARTIQEELERMAFDDRP EKORIFHRTTHIEMIKGEFCCKGIENYSRHFTGAPPGAPPCLLDYFPEDFLIIDE SHQTLQIRAMVGRQSRKQSLVEYGFPLPSAFDNRLPYVEEAQYFRKYIYVSATPG DVTQESSGHIYVQOIIIRPTGIPDPMPETIRPTQITDLRSQVTDVILGVNLLREGDIL TKKLADMGAFVISELIEPAAYLHSGIETARTQITDLRSQVTDVILGVNLLREGDIL PEYSLVALADAKEGFLRSTSLQFCGCAARNINCKVIFYADOKTRTSIEETLRETER RRQIQLDYKNEINVPKTIKAIIPANPILOTSKDSPEKSPQRSKLEDEQIKKYE ALMORAKREFRENAKYRDMOACKEQLLYLF" complement(7327)..8361 /gene="trpS" complement(7327)..8361 /gene="trpS" /codon_start=1	gene CDS	1889.00	alignment_scores:	Quality: 362
gene CDS	1889.00 /transl_table=11 /product="tryptophanyl tRNA synthetase" /protein_id="BAA99010.1" /db_xref="GI:8979176" /translation="MNKKRVLTGDRPTGKHLGHVWGSIKNRLEQNSPEYDCFFII ADLHTLTTRKEEVDNDHIVEVLADWLSVGDPTKSIYIYQSAIPEIYELHLDFS MLISINVMGIPSLKDMARNASIEGSLSYGLIGYPILOQSAIILAKAOPVPVKGONE AHVELTDIARNRNLYGOVPEPEVLOGELTSLVGDGCKMSKSNANAIYLSDSDA TITEKVTMYTDNRIARTPGCEVGNPLFIYHDIENPHKDEVEEPKARYOGCIGDI EVKARLAEBLILHLKFIKERRSEFLSKPLALQNVLEDGTHMKREYAKVYTHEEYHDKFG FSHKWSLLK" complement(8370)..8924 /gene="CPJ0803" complement(8370)..8924 /gene="CPJ0803" /codon_start=1 /transl_table=11 /product="ct584 hypothetical protein" /protein_id="BAA99011.1" /db_xref="GI:8979177" /translation="MAAKTKTLEEDNVFLLEGNLKRIFATPIGYTTFREFONVVFN CANGQBIANFFEFMEIKNGKLQELAPQOQAASHLIAEFMMPIRVAKDITHEREGFIIN FITSDMLTQERCIFLNRILARVDGQFELMTDVQNTCHLIRHLARLEAKKNPVGEK NLQIQEITSLKHFDELTALQ" complement(8928)..9680 /gene="gp6D" complement(8928)..9680 /gene="gp6D" /codon_start=1 /transl_table=11 /product="CHLTPR plasmid paralog" /protein_id="BAA99012.1" /db_xref="GI:8979178" /translation="MGNLATLLESKPKNTPTKMEALARKMEGDPSPFLAVRLSNPTL SSKEQIQRHLLOHYNFRSQIEEPDLOICLSAEVQIHHQSVLLHGERITKVRDIL KSYREGAFSSWLLTYGNQTYNFIYELFTLLPELTKITEMEKMPQRAVYTLASRQ GPOKKEEIRNRYGERKSELLDRRIYFELVETDCRKTSPVQOALAMTKGSQLTK CTSLSDQEIIILKIKLEKVKSNLFFPDTKV" complement(9685)..10452 /gene="mind" complement(9685)..10452 /gene="mind" /codon_start=1 /transl_table=11 /product="chromosome partitioning ATPase" /protein_id="BAA99013.1" /db_xref="GI:8979179" /translation="MKTIAVNSFKGSGTAKTTLHLGAALAOYHOARVLILIDFQACAN LTSGLGDPDCYDSLAVVLOGEKEIQEVIRPIQDTOLDLIPADTMLERIEVSGNLAAD RYSHERKLYVSGVDKDYVILIDTPPSICWLTSAIADYALICATPEYSVKGLE PLAGIQCISARPLILGYVALSWNCCKNSAFELIHKTFPGKLLNTRKIRRDITV SEARHGKVPFATSPSARASEDYFNLTAKELLILLADI" complement(10485)..12392 /gene="thrS" complement(10485)..12392 /gene="thrS" /codon_start=1 /transl_table=11 /product="threonyl tRNA synthetase" /protein_id="BAA99014.1" /db_xref="GI:8979180" /translation="MIQVTCQDNKYEVLEGTAAELAKOLKNSHQFIVLINERPRDL STHLNEGDTLVFTIETSDPSEIRFIHTSAHLLAQAVLRWLPDAITPTIGPVYDHGFYD FANLISESDVFIETDVQIYDEKLAISRETYGDKQOALAQFPQNPFTELIRELPE NEETSAYSQGEFDICRPHLPSTAHKAFKVLRTSAAYWRGDPRESLRYIYGTSEFP TSKELRAHLQIEBAEKRDHVRVLGAKLDLFSQESSSPMPFFPHPRGMVMDALIRYWK QHTAACYKEILPOLMNRQLMEVSGHDNYKANNYTLQIDDEYATKPMNCPGCMY YKTLRHSYGLTGFLEHYHLELSTREPKDTIGDLSLWELATDALNRALVOSGTFPIVRP NILQVNSTLYGTGLEYHLELSTREPKDTIGDLSLWELATDALNRALVOSGTFPIVRP GEGAFYGPKIDLIHVKDAIQRTWQCCTIOLDMFLPERFELEYTTAOGTKSVPMHLRAL	gene CDS	1889.00	alignment_scores:	Quality: 362

Ratio: 5.218  
Percent Similarity: 100.000  
Percent Identity: 100.000  
Gaps: 0  
alignment\_block:  
US-09-438-185-1047 x AP002548/rev ..  
Align seg 1/1 to reverse of: AP002548 from: 1 to: 325865  
1 ValHisTyrCysGluArgThrLeuAspProLysTyrIleLeuLysIleAl 17  
296378 GTGCACCTACTGCGAGAGAACCCCTGGACCCAAAGTATATCTCGAAGATTGC 296329  
17 aleuLysLeuArgGlnSerLeuSerLeuPheGlnAsnSerGlnSerL 34  
296328 TCTAAAGCTGAGACAATCACTTTCCCTGCTTCTCCAGAACAGCCAAATCAC 296279  
34 euGlnArgAlaTyrSerThrProTyrSerTyrTyrArgIleIleLeuGln 50  
296278 TCCAACTGTGCATACTCGACCCCATATCTCTACTACCGAATCAATCTACAA 296229  
51 LysGlnAsnLysGluLysGlnAlaLeuAlaArgHisLysCysIleSerL 67  
296228 AAGGAAATAAAGAGAGAACGCTTAGCTGCACACAAAGCAATTCAT 296179  
67 eLeuGluPhePheLysAsnLeuLeuPheValHisLeuLeuSerLeuSerL 84  
296178 TTTAGAAATTTTCAAAACATCTCTTTGTTTCATCTTCTGTGCATATACAA 296129  
84 YAsnGlnArgGluGlyCysSerThrAspMetAlaValValSerThrPro 100  
296128 AGAATCAAAAGGAAGGTGCTCCACATGATATGCTGTGTGTAAGCACTCCC 296079  
101 PhePheAsnArgAsnLeuTyrTyrArgLeuLeuSerSerArgPheSerLe 117  
296078 TTTTAAATCGGAATTTATGCTATGCTACGACCTCTTCTCCACGGTTTCTCT 296029  
117 utrLysSerTyrCysProArgPhePheLeuAspTyrLeuGluAlaPheG 134  
296028 ATGGAAGAGCTATTGTCCAAAGATTTTCTGATTTACTTAGAAGCTTTCG 295979  
134 lLeuLeuSerAspPheLeuAspHisGlnAlaValIleLysPhePheGlu 150  
295978 GCTCTCTTCGATTTCTTAGACCATCAAGCATTAATTTCTCGAA 295929  
151 LeuGluThrHisPheSerTyrTyrProValSerGlyPheValAlaProHi 167  
295928 TTAGAACAACATTTTCTATATFCCGGTTTCAGGATTTGAGCTCCCA 295879  
167 sGlnTyrLeuSerLeuLeuGlnAspArgTyrPheProIleAlaSerValM 184  
295878 TCAATACTGTCTCTGTGTCAGGACCGTACTTCCCTTCCCTGCTGTAA 295829  
184 eArgThrLeuAspLysAspAsnPheSerLeuThrProAspLeuIleHis 200  
295828 TGGCAACTCTCATTAAGATAATTTCTCTTACTCTGATCTCATCCAT 295779  
201 AspLeuLeuGlyHisValProTyrPheLeuHisProSerPheSerGluPh 217  
295778 GACCTTTTAGGCACGCTGCCCTTGGCTTCTACATCCCTCATTTTCTGAAT 295729  
217 ePheIleAsnMetGlyArgLeuPheThrLysValIleGluLysValGlnA 234  
295728 TTTTCATAACATGGGAGACCTCTTCACTAAAGTCATAGAAAAAGTACAA 295679  
234 laLeuProSerLysGlnArgIleGlnThrLeuGlnSerAsnLeuIle 250  
295678 CTCTTCTAGTAAACACGACCATCAACCCCTCAAGCAATCTGATC 295629  
251 AlaIleValArgCysPheThrPheThrValGluSerGlyLeuIleGluAs 267  
295628 GCATTTGAGCTGCTTTTGGTTTACTGTTGTAAGCGGACTTATTGAANA 295579  
267 nHisGluArgLysAlaTyrGlyAlaValLeuIleSerProGlnG 284

295578 CCATGAAGGAAGAAAGCATATGAGCGGTTCTTATCAGTTCTCCTCAGG 295529  
284 luLeuGlyHisAlaPheIleAspAsnValArgValLeuProLeuGluLeu 300  
295528 AATTTGACACGCTTTTCATTGATAACGTACGTGTTCTCCCTTTAGAATTG 295479  
301 AspGlnIleIleArgLeuProPheAsnThrSerThrProGlnGluThrIle 317  
295478 GATCAGATATTCGCTCTCCCTTCATACATCACTCACTCCACAGACACTT 295429  
317 uPheSerIleArgHisPheAspGluLeuValGluLeuThrSerLysLeuG 334  
295428 ATTTTCAATAAGACATTTTGTATGAACTGGTAGAACCTCCTCAAAATTAG 295379  
334 luTrpMetLeuAspGlnGlyLeuLeuGluSerIleProLeuTyrAsnGln 350  
295378 AATGGATGCTCGACCAAGGCTCTGTAGAAATCAATTCCTCTTACAATCAA 295329  
351 GluLysTyrLeuSerGlyPheGluValLeuLeuCysGln 362  
295328 GAGAAATATCTTCTGCTTTGAGGTACTTTGCCAA 295293  
seq\_name: gb\_bal:AE004410  
seq\_documentation\_block:  
LOCUS AE004410 10143 bp DNA BCT 31-JUL-2000  
DEFINITION Vibrio cholerae chromosome II, section 67 of 93 of the complete  
chromosome  
ACCESSION AE004410 AE003853  
VERSION AE004410.1 GI:9658244  
KEYWORDS  
SOURCE  
ORGANISM  
Vibrio cholerae  
Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.  
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,  
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,  
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,  
Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragol, I.,  
Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,  
Nierman, W.C. and White, O.  
DNA sequence of both chromosomes of the cholera pathogen Vibrio  
cholerae  
NATURE 406 (6795), 477-483 (2000)  
JOURNAL  
MEDLINE  
20406833  
2 (bases 1 to 10143)  
REFERENCE  
AUTHORS  
Heidelberg, J.F., Eisen, J.A., Nelson, W.C., Clayton, R.A., Gwinn, M.L.,  
Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Umayam, L.A.,  
Gill, S.R., Nelson, K.E., Read, T.D., Tettelin, H., Richardson, D.,  
Ermolaeva, M.D., Vamathevan, J., Bass, S., Qin, H., Dragol, I.,  
Sellers, P., McDonald, L., Utterback, T., Fleischmann, R.D.,  
Nierman, W.C. White, O., Salzberg, S.L., Smith, H.O., Colwell, R.R.,  
Nekalanos, J.J., Venter, J.C. and Fraser, C.M.  
Direct Submission  
Submitted (14-JUN-2000) The Institute for Genomic Research, 9712  
Medical Center Dr, Rockville, MD 20850, USA  
FEATURES  
source  
1. 10143  
/organism="Vibrio cholerae"  
/strain="N16961"  
/serotype="O1"  
/db\_xref="taxon:666"  
/chromosome="II"  
/note="biotype: El Tor"  
97. . 681  
/gene="VCA0817"  
97. . 681  
/gene="VCA0817"  
/note="identified by Glimmer2: putative"  
/codon\_start=1  
/transl\_table=11  
/product="hypothetical protein"



```

/protein_id="AAF96715.1"
/db_xref="GI:9658245"
/translation="MSITVRYTMEYLLALLOTHYKWLMTILIIIVLYPHISRTIRLRFES
MLTRGVNDIHLKRAKRLIRTLILFVILWFTTILWGLIELGLLWVGSLSFALIGVGLFA
GWSLSNVTAFLILFVQNCRIGCWVRIVDCANFIEGKIIDMALLNVELETLEGNRLV
YPNNLFTVRPVIVLQKAPEKAVKPERKLPKGL"
gene
CDS
/genes="VCA0818"
/genes="VCA0818"
/genes="VCA0818"
/genes="VCA0818"
/notes="similar to PID:1652477; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="magnesium transporter MgtE, putative"
/protein_id="AAF96716.1"
/db_xref="GI:9658246"
/translation="MSEYQDLSQQIELIATAESEQTHLIGEMVENGDEGSIALILE
AFPEQVRIMRSLPLEMHIDVLTQMAEVRISVINALSIEKLT/LAKLDNLSLIEW
EDSLPDIIESEALSINRDELELYDQAEADDEIGHWADRVKVTYLPFISVVRRAKVL
LDKYHYDSQYIYVILNKAKRGLVFSFSDVLAQETQLKNRMEELVTINAKQPTLSD
AVDALHTLPIVPLVSDDETLEIGDHFALSQVREMYEARLMAGTGMDEGDDLPAP
VIKSKKRGIMLGINLLTALASITIGLFENVISQVVALVLMPIVASMGGIAGSOTL
TLVVRAMALNOITVGNRFALLKNELGIGSINGLWALIGGMAALVQFQEPVIGATIAL
AIVVNITTAALFGLPIPLIDDKFLPALAGSVILITGDCRGLFCLPRHROFGDVA
AIRP"
gene
CDS
/genes="VCA0819"
/genes="VCA0819"
/genes="VCA0819"
/notes="similar to SP:P05380 GB:X07850 GB:X07899 PID:41613
PID:146269; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="chaperonin, 10 kd subunit"
/protein_id="AAF96717.1"
/db_xref="GI:9658247"
/translation="WNIRPLHDKLIVERLEVENKSEGGIVLTSSQVSKSNRGRKVAVG
IGRLPKNGDRARMEVKTGDIIFNGYGVTEKVDGKEXYLILSESVDVAIV"
gene
CDS
/genes="VCA0820"
/genes="VCA0820"
/genes="VCA0820"
/notes="similar to GB:U14003 SP:P06139 GB:X07850 GB:X07899
PID:146269; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="chaperonin, 60 kd subunit"
/protein_id="AAF96718.1"
/db_xref="GI:9658248"
/translation="NAAKHLVFSTDAQKMLSGVNLNANAVKVLGPKGRHVVLNKSX
GAPTIDGVSVAKELELADKFERMGQMLKQVASKANDEAGDGTATVLAQALINE
GMKAVANGNPNMOLKRGIDKAVSANEXKLHQLAKPCSDTQSIQVGAISANSRHAIGE
IIOAKMEKVGNGYITVEEQNALQNELSVVEGQFDGRLSPFFINQPKSGVLEENP
YVLVDKIKSHIRLLELPILLESVAKASRLLIIFEDVDGDALATLVVNSHRGLIKVAAV
KAFGEQRKMELEALITAGRVISIEIGLELEKLVTLDDLGSAKVTINKDNTTIVD
GAEPSSALQDRTAQIQHQLTTSQYDHDKLIQAKLSGGVAVIKIGATAEVENKEK
KDRVDALHATRAAEVGEIVAGGVALIKTANELSLQGDNDQNVGTRIALRAMEEP
LRQIATNAGDEASVIANOVKTGDHYGYNAATGQFNGNMLEMGILDPKAVTRSAIQFAA
SIAGLMITTEAMVSEVDKESAA"
gene
CDS
/genes="VCA0821"
/genes="VCA0821"
/genes="VCA0821"
/notes="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF96719.1"
/db_xref="GI:9658249"
/translation="WTKLPHQLSVVEAQPKTHPANAPMNSFPQRCODRVLMPPYR
SVSPAVLVVYVYINQAOQPSPTLH"
complement(4464..5888)

```

```

/gene="VCA0822"
complement(4464..5888)
/gene="VCA0822"
/notes="similar to GB:AE000782; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="aspartokinase, putative"
/protein_id="AAF96720.1"
/db_xref="GI:9658250"
/translation="MMSYTVKEIGCTSMTADAVLENILLPEOLYGRVTVASXSGM
TDALLECKRPGAGIYQIKIAKHDESWRDVAYALEQRMILLINEMFADMTLRADKFI
RARLARIACIQNILQOTQGFSLRQIPQIREFPLASLGESNAYNTALKNLISYN
ARFVDSLGCWCEANGLNDQVIEEAFRDIDVTSELPVITVYASQCEGLMRTIDRGISEM
TFSRIAVVTAQQAIIHKETHLSSADPRLVGAEKVRPIGTUNFEDVADQLANIGWEAIIH
PNAAGLRVQSNIDILRIKTFEPDHSGLTITASKEMANSVTYYLSGSEPLNHRIVQAEKRYPOA
MVGMDKVSHELIEHIAAKVOLIGKEMANSVTYYLSGSEPLNHRIVQAEKRYPOA
ANGCVLHQALMPFAQPSARKVA"
complement(5885..6301)
/gene="VCA0823"
complement(5885..6301)
/gene="VCA0823"
/notes="similar to PID:2098610; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ectoine synthase"
/protein_id="AAF96721.1"
/db_xref="GI:9658251"
/translation="MIYRLIECROSERRVVAENWESVRMLLKKDDHMGFSFHTITIYA
NTQTHIHRNLESVYCMGSEIEVYGVGKTYIPQPGTYLIDQHDHDLHAFSSSEM
MACVFNPLTGRHEIHQAEVYPLDKSELISQCHREK"
complement(6366..7601)
/gene="VCA0824"
complement(6366..7601)
/gene="VCA0824"
/notes="similar to PID:2098609; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="diaminobutyrate--pyruvate aminotransferase"
/protein_id="AAF96722.1"
/db_xref="GI:9658252"
/translation="MDIFKHHSQVQSVYANHFVPLFGTAKGSWLYSQGDVAYLDFLSG
AGALNYGHNAVLKQALLEYIERDGLTHGLDMHSEAKAHFTQALQTHILEPRLNYKL
OFTGPTGTNAVEAALKARKVTGRHNVYFTNGFHGCSLGAATGNHHRQAGIAL
GVYVVPDGYAGVDGLTLFETMLDNSSGLDKPAVLLLETVOGEGNLVNSADWLQR
VQAICRAQOILLIVDDIQAGCGRTCTFFSPSGIIEPDMVTLSKLSGLGVLPMALVLF
KPEWDMQKPEHNGTFRGNHAFVTATRALEYANODFOTLIAARSEQVTAQLQCL
SRVPTLSEGLKGLMOGLACHNGDIARDIAALCFQKGLLIETNAGEDEVILKVFCELT
ITENDLAHGLTIIEVVLLS"
complement(7625..8146)
/gene="VCA0825"
complement(7625..8146)
/gene="VCA0825"
/notes="similar to SP:052249; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="L-2,4-diaminobutyric acid acetyltransferase"
/protein_id="AAF96723.1"
/db_xref="GI:9658253"
/translation="MIYPOIMHKPALPWVRRPRTQEDGLSIHELIAQCAPLQONSAYC
NFLQSSHFOYTCMLMAEQELVGVFSAYRKPQEQNELFIWQVAVHPSARGKGLAYQML
KHILAREDLADITVLETTITRSNQASWRLFKOKLDREOQEGSVSTFLDETCHFEGEHD
TEYLYRIPIQSSN"
complement(8259..8375)
/gene="VCA0826"
complement(8259..8375)
/gene="VCA0826"
/notes="identified by Glimmer2; putative"
/codon_start=1

```



/transl\_table=11  
/product="hypothetical protein"  
/protein\_id="AAF96724.1"  
/db\_xref="GI:9658254"  
/translation="MLIRELAAYQPHRYFKIHFMFOFKVFLFLAKGFFSF"  
complement(8846..9175)  
/gene="VCA0827"

alignment\_scores:  
Quality: 256.50 Length: 254  
Ratio: 1.666 Gaps: 6  
Percent Similarity: 60.630 Percent Identity: 25.591  
alignment\_block:  
US-09-438-185-1047 x AE004410/rev ..  
Align seg 1/1 to reverse of: AE004410 from: 1 to: 10143  
106 LeuTrpTyrArgLeuLeuSerArgPheSerLeuTrpLysSerTyCy 122  
9890 GTTTGGCATGAGTATCACTCCGCGCAAGAGTGTGAAACACGAGC 9841  
122 sProArgPheLeuAspTyrLeuGluAlaPheGlyLeuLeuSerAsp 139  
9840 CTGCCAAGCTTATTAGATGGTTGAATATGCTTAATTTGCCGACAGATC 9791  
139 heLeuAspPheGlnAlaValIle..LysPhePheGluLeuGluThrHis 154  
9790 GCCTGCCACAATGCCGGAATATAAACCAGGATTTGCAAGAGAAACCGGT 9741  
155 PheSerTyrTrpValSerGlyPheValAlaProHisGlnTyrLeuSe 171  
9740 TGGCAAGTTGAGCTGCTCGGCATTGATGATGCTTTCATGCTTCTTGC 9691  
171 rLeuLeuGlnAspArgTyrPheProIleAlaSerValMetArgThrLeu 188  
9690 CTTACTCGCCGATAAGAAGTTTCCAGTGCCACGCTTTTTCGGCGTAGAG 9641  
188 sPLysAspAsnPheSerLeuThrProAspLeuIleHisAspLeuGly 204  
9640 AGAGTTTGATATCTCCAGAGCCCGATTTCTCCATGAGGTATGTT 9591  
205 HisValProTrpLeuHisProSerPheSerGluPhePheIleAsnMe 221  
9590 CACTCGCTATGCTCAGCCACCCCTGATTTGCGCTTTTACTCATGTTA 9541  
221 tGlyArgLeuPheThrLysValIleGluLysValGlnAlaLeuProSerL 238  
9540 TGGCCAGCTAGTGTCTAAAGCCACCCCAAGAGCGTAGCTAC..... 9498  
238 ysLysGlnArgIleGlnThrLeuGlnSerAsnLeuIleAlaIleValArg 254  
9497 .....TTAGCCGCT 9489  
255 CysPheTrpPheThrValGluSerGlyLeuIleGluAsnHisGluGlyAr 271  
9488 CTGATGTTTACGGTTGAGTTGCGTTAGTTTCAAGAACAGGACAAAC 9439  
271 gLysAlaTyrGlyAlaValLeuIleSerSerProGlnGluLeuGlyHisA 288  
9438 CAAATTTATCGCGGGGCGATTTTGTGCTCAGCCGCTGAACAGCTTTTACG 9389  
288 laphHeileasPasnVal...ArgValLeuProLeuGluLeuAspClnile 303  
9388 CTCAGAAAGCACATCCAAACCGTCAGCCGCTCGATATTATGCAAGTG 9339  
304 lIeArgLeuProPheAsnThrSerThrProGlnGluThrLeuPheSerIl 320  
9338 CTACGACCTCTTACCGATGATGATCATGACCGGATTTATTATGCTT 9289  
320 eaArgHisPheAspGluLeuValGluLeuThrSerLys.....L 333  
9288 GCGGATCTCTAGTCAACTGATCAACTCAGTCAACGCGATGTAATGGCGT 9239

333 euGluTrpMet...LeuAspGlnGlyLeuLeuGluSerIleProLeuTyr 348  
9238 TGGTGTGCAAGCTATGCAAGACGAGTACTCCG.....CCTCTTTT 9195  
349 AsnGlnGluLys 352  
9194 CAACCAAGGAA 9183  
seq\_name: gb\_ov:GGA251387  
seq\_documentation\_block:  
LOCUS GGA251387 1633 bp mRNA VFT 02-DEC-1999  
DEFINITION Gallus gallus mRNA for tyrosine hydroxylase (tyrosine hydroxylase gene).  
ACCESSION AJ251387 GI:6523292  
VERSION AJ251387.1  
KEYWORDS tyrosine hydroxylase; tyrosine hydroxylase gene.  
SOURCE chicken.  
ORGANISM Gallus gallus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.  
REFERENCE 1 (bases 1 to 1633)  
AUTHORS Ernsberger, U., Patzke, H., Tissier-Seta, J.P., Reh, T., Goridis, C. and Rohrer, H.  
TITLE The expression of tyrosine hydroxylase and the transcription factors c-fos-2 and c-myc-1: evidence for distinct inductive steps in the differentiation of chick sympathetic precursor cells Mech. Dev. 52 (1), 125-136 (1995)  
JOURNAL 96076133  
MEDLINE  
REFERENCE 2 (bases 1 to 1633)  
AUTHORS Ernsberger, U.  
TITLE Direct Submission  
JOURNAL Submitted (01-DEC-1999) Ernsberger U., Ruprecht-Karls-University, Institut f. Anatomie und Zellbiologie III, Im Neuenheimer Feld 307, D-69120 Heidelberg, GERMANY  
FEATURES  
source  
1..1633  
/organism="Gallus gallus"  
/db\_xref="taxon:9031"  
/dev\_stage="embryonic day 8"  
/tissue\_lib="sympathetic ganglia"  
32..1507  
/gene="tyrosine hydroxylase"  
32..1507  
/gene="tyrosine hydroxylase"  
/codon\_start=1  
/product="tyrosine hydroxylase"  
/protein\_id="CA862388.1"  
/db\_xref="GI:6523292"  
/translation="MPTPNISTSAKFRRAVSELDKQAEATMSPRFICRRQSLIED ARKREAAAAATAESTETIVFEKDGRLNLFMLKGAKTSPLSRAKLVFEFEFA KIHLETRLSKRPGETABLEVFEVCEVHSDLNFTISSIKRVAEDVTRTKEDKHFHF PRKICELDKCHHLVTKFDLPDLDPGYSDOVYRQRKSAIEAFYKHGDPTRPVEY TAETATWKEVYSTLKSLYTHACKYLEAFENLLEKFCGYNNENIPOLEEVSRLPKER TGFQLRPVAGLLSARDFLASLAFRVFOCTOYLIRHASSPMHSPEDCCHELLGHVPLA DKTFASODIGLASCATDDETEKLATLYMTVEFGLCRQNGIVKAYGALLSSVGE LHSLSDEPEVRDFPDMAAVQPDONQVQPVFVSFSDAKNLRNYAAHIKRPFS VKYEPYTHSIELSDSPOTICHSLEVRHLSLINALNVIS"  
BASE COUNT 412 a 448 c 405 g 368 t  
ORIGIN  
alignment\_scores:  
Quality: 237.50 Length: 227  
Ratio: 1.827 Gaps: 4  
Percent Similarity: 57.269 Percent Identity: 29.075  
alignment\_block:  
US-09-438-185-1047 x GGA251387 ..  
Align seg 1/1 to: 1633



```

181 TCATCTCCCAAGAAATGGC.....:
232 alGlnAlaLeuProSerLysGlnArgIleGlnThrLeuGlnSerAsn 248
1004 TGGCTCTCTTGGAGCTTCAGAGGACGGTTCAGAACTGGCAACG... 1050
249 LeuIleAlaIleValArgCysPheThrPheThrValGluSerGlyLeuIle 265
1051 .....TGCTACTCTCTTCTTCTGAGTGTGGACTGTG 1082
265 eGluAsnHisGluGlyArgGlyAlaValGlyAlaValLeuIleSerSerP 282
1083 CAGCAGATGGCCAGCAGGAGCTTGTGGCGGCTGCTTCTTCCA 1132
282 roGlnGluLeuGlyHisAla.PheIleAsnValArgValLeuProIle 298
1133 TCAGTGAGCTCAGACATGACCTTCTGGACATGCCAGGTTAGGCCCTT 1182
298 uGluLeuAspGlnIleIleArgLeuProPheAsnThrSerThrProGlnG 315
1183 G.....ATCCCAAGTTCCTGCAACACAGGAATGCTCATCAC 1220
315 lu 315
1221 ka 1222

seq_name: gb_bai:AE004522

seq_documentation_block:
LOCUS AE004522 15724 bp DNA BCT 30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 83 of 529 of the complete
genome
ACCESSION AE004522 AE004091
VERSION AE004522.1 GI:9946768
KEYWORDS
SOURCE Pseudomonas aeruginosa.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
1 (bases 1 to 15724)
Stover,C.K., Phan,X.Q., Erwin,A.L., Mizoguchi,S.D., Warren,P.,
Hickey,M.J., Brinkman,F.S., Hufnagle,W.O., Kowalik,D.J., Lagrou,M.,
Garber,R.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Larbig,K., Lim,R.,
Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
NATURE 406 (6799), 959-964 (2000)
20437337
2 (bases 1 to 15724)
Stover,C.K., Phan,X.Q., Erwin,A.L., Mizoguchi,S.D., Warren,P.,
Hickey,M.J., Brinkman,F.S.L., Hufnagle,W.O., Kowalik,D.J.,
Lagrou,M., Garber,R.L., Coulter,S.N., Folger,K.R., Kas,A.,
Larbig,K.R., Kas,A., Larbig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K., Wu,Z., Paulsen,I.T., Reizer,J., Saier,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
FEATURES
source
location/Qualifiers
1..15724
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
complement(188..1387)
/gene="phhC"
/notes="PA0870"
complement(188..1387)
/gene="phhC"
/codon_start=1

gene
complement(188..1387)
/gene="phhC"
/codon_start=1

CDS
complement(188..1387)
/gene="phhC"
/codon_start=1
```

```

/transl_table=11
/product="aromatic amino acid aminotransferase"
/protein_id="AAG04259.1"
/db_xref="GI:9946769"
/translation="MSHFAYKAVRPGDPTILGLLDVARNPDAKLDLGVGVYKDAQGL
TPILSYLAGRLVYQETPKSYVGCGHGDALFAARLAEALGAASPLLEQRADATCT
PGTGALLRAGDFLAHCLPAGCIWSDPTWPIHETLEAAGLKVSHYPTYSADNRLDV
EAMLAGLERIPQGVYVLLHACCHNPFGEDLSHDDMFVLDVVRRELLPLIDPAYQGF
GGLEEDAWAVRLFAGELPEVITSSCSNFGLYRDRVGCALIVCAONAKILTLDRSLG
AFLARHMTSTPAGAEVVAAILGDSLKGMOEVEGMRSLASRLGLVLEALAPG
LAERFAHVGQGRGMSYTGSLPQQVAVRLRDEHAHVLVSSGRANVAVGLDRRLDLAQA
IAQVCAD"
complement(1384..1740)
/gene="phhB"
/notes="PA0871"
complement(1384..1740)
/gene="phhB"
/codon_start=1
/transl_table=11
/product="pterin-4-alpha-carbinolamine dehydratase"
/protein_id="AAG04260.1"
/db_xref="GI:9946770"
/translation="NTALTOAHCEACRADAPHVSDLEPVLRLQIPDWNIEVRDGIQ
LEKYLKPKHALAFTNAVGEISEAGHHPGLLTWGVKVTVTWWSHSIKLGRNDFI
MAARTDEVAKTAEGRK"
complement(1845..2633)
/gene="phhA"
/notes="PA0872"
complement(1845..2633)
/gene="phhA"
/codon_start=1
/transl_table=11
/product="phenylalanine-4-hydroxylase"
/protein_id="AAG04261.1"
/db_xref="GI:9946771"
/translation="MKTQYVARQDDNGFTHYPETEHWVNTLITRQLKLVIEGRACQ
EYLDGIBOLGHPHERIPQDEINRVLOATTWGRVAVPALIPQTFELLASQQPVA
TFITPELDYLOEPDIFHEIFGCHPLTNPFWAEFTHTYKLGKLSKEERVFARL
YMTFIEPLVETDOGRYVGGISLPKRYTYSLSDEPLHQAFLNPLEAMRTPTRIDIL
QPLYFVLRLDLKRLFLQAQEDINALVHEAMRLGHAPLFPFKQAA"
2917..4476
/gene="phhK"
/notes="PA0873"
2917..4476
/gene="phhK"
/codon_start=1
/transl_table=11
/product="transcriptional regulator PhhK"
/protein_id="AAG04262.1"
/db_xref="GI:9946772"
/translation="MRKVKQNRVGIILRDITLNLVDYGVNVRGEGVGGDGNATVLL
CPNMINLOLSGLRPLENPGVGVKRVGLMPSEHLELALLAALDPVLSDVMGQ
GIVANRAAQLGRLVNDVPIPIRVRVOLDLPELVANRANGLRKYKVGDMQV
ADIALQSEHDESEALAGAVLILHRADRVGRIVYRQELRGFDSIFQSSRMAAVV
REARMAPLADPLIIEGTGTGKELLARACHLASRQSPFALNCAGLPGESMAFEL
FYGYGFEGARPEGKIGLLLEITAGTFLDGVGEMSPRLOAKLILPDGCGFRVGS
DEEYLDVRYCATQVDLSLCAKGEFRODLTHRLNLSLHITPLIECLDGLAPLAEH
FLDQASQIGCGIPKLSQAQALKERLRYHWPGRQLENLFPQAVSLCEGCTYKAHIR
LPDYGAQPIGDFSLGLEDLDAIVGRFKAVERLFRHPSRQLGRKRLGVSHHTAANK
LRQGVGQSEG"
complement(4682..4948)
/gene="PA0874"
complement(4682..4948)
/gene="PA0874"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAG04263.1"
/db_xref="GI:9946773"
/translation="MKAFALMALFACSLPVMAGAQDQFASQQENAVEEYAYGMQLDV
ARVISSTDTSNACGIVFARMTYEDSNQKHTVEYQVWNGNGCTG"
complement(5217..7412)
/gene="PA0875"
complement(5217..7412)
/gene="PA0875"
```





```

MEDLINE 97072811
REFERENCE 2 (bases 1 to 1575)
AUTHORS Florez.J.C., Seidenman.K.J., Barrett.R.K., Sangoram.A.M. and
Takahashi.J.S.
TITLE Direct Submission
JOURNAL Submitted (05-MAY-1995) Neurobiology/Physiology, Northwestern
University, 2153 N. Campus Drive, Evanston, IL 60208, USA
FEATURES
source
1. 1575
/organism="Gallus gallus"
/strain="white leghorn"
/sub_species="domesticus"
/db_xref="taxon:9031"
/tissue_type="pineal gland"
/clone="pTPH-37"
1. 82
/EC_number="1.14.16.4"
/codon_start=1
/product="tryptophan hydroxylase"
/protein_id="AAC60036.1"
/db_xref="GI:3645921"
/translation="MIEDNKKNDHAPGRGTAIIFSLKNEVGLVKALKLFQEKHVN
LVHIESKRRNRSEFIVDCDSNRQLEIFQLKSHSVISVMNPTHEHFNQEDGD
MENIPWPKKISDLDCANRVLMDGDLADHPGKDNVYKRKYFADLANNYKHGD
PIPIETEEIKTWGTVRELKLYPHACREYKLNPLITRYCGYREDNIPQLEDV
SRFLKRTGPTIRPVAGYLSPRDLAFLAFVHCTOYVRHSSDPLYTPEPDTCHLL
GHVPLAEPFAQSEIGLASLGADEAVOKLATCYFTVFEGFLCKQSGOLRVYAG
LLSISLKHLSGSAKVPDPKVTQOECLITTFQEVYFVSESEFEAKRMREFAK
TIKRPFGKYNPTQSVQILDKSIASVNNELRHELDIVSDLSKMGKQLEV"
1421. >1575
3'UTR 433 a 368 c 381 g 393 t
ORIGIN

alignment_scores:
Quality: 231.00 Length: 229
Ratio: 1.819 Gaps: 4
Percent Similarity: 55.459 Percent Identity: 28.384

alignment_block:
US-09-438-185-1047 x GGU26428

Align seg 1/1 to: GGU26428 from: 1 to: 1575

108 TyrArgLeuLeuSerSerArgPheSerLeuTriplysSerTyrCysProAr 124
||||| |||||
617 TACCGAGAGCTTAAC .....AAGCTTTACCCCACTCA 648

124 gPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheLeuA 141
||||| |||||
649 TGCGTCGAGAGAGTACTTAAAACTTCCCTTGCCTCACCANAATCTGTG 698

141 sPHis .....GlnAlaValIleLysPhePhe 149
||||| |||||
699 GTTACAGGGAAGACAAATATCCCCAGCTGGGAAGAGTGTCCTCCGTCCTG 748

150 GluLeuGluThrHisPheSerTyrTyrProValserGlyPheValAlaPr 166
||||| |||||
749 AAAGCGCGACAGGTTTCAACATTCGCCAGTCCTCGCTATCTGTCAAC 798

166 oHisGlnTyrLeuSerLeuLeuGlnAspArgTyrPheProIleAlaSerV 183
||||| |||||
799 CAGAGACTCTTGGCAGGATAGCATTACAGAGTTTTCACCTGCACCTCAGT 848

183 alMetArgThrLeuAspLysAspAsnPheSerLeuThrProAspLeuIle 199
||||| |||||
849 ATGTTAGACACAGCTCGGACCTCTCTACACACAGACAGCTGATACCTGC 898

200 HisAspLeuLeuGlyHisValProTyrPheLeuHisProSerPheSerG 216
||||| |||||
899 CATGAGCTCTAGGCAATGCTCCTCTTTTGCTGAACCCAGCTTTTGCTCA 948
```

```

216 uPhePheIleAsnMetGlyArgLeuPheThrLysValIleGluLysValG 233
||||| |||||
949 GTTCCCGAGAAATGGT .....CTTG 971

233 InAlaLeuProSerLysLysGlnArgIleGlnThrLeuGlnSerAsnLeu 249
||||| |||||
972 CATCATTGGCGCATCAGAGCGGTGTCACAAAACCTGGCAACA..... 1015

250 IleAlaIleValArgCysPheThrPheThrValGluSerGlyLeuIleG 266
||||| |||||
1016 .....TGCTACTTTTTCACGTAGAGTTTGCCCTGTGCAA 1050

266 uAsnHisGluClyArgLysAlaTyrGlyAlaValLeuIleSerSerProG 283
||||| |||||
1051 CCAAGAGGACAGCTTAGAGTTTATGGGGCTGCTCTCTCTCTATTA 1100

283 InGluLeuGlyHisAlaPheIleAspAsnValArgValLeuProLeuGlu 299
||||| |||||
1101 GTGAGCTCAAGCACTCGCTCTCTGCGAGTGCCAAAGTCAAGCCTTTTGAT 1150

300 LeuAspGlnIleIleArgLeuProPheAsnThrSerThrProGlnGluTh 316
||||| |||||
1151 CCAAGGTCACCTGCAAGCAAGAATGCTGTATTACAACTTTCCAGAGGT 1200

316 rLeuPheSerIleArgHisPheAspGluLeuValGlu 328
||||| |||||
1201 TTACTTTGTTCTGAAAGTTTGAAGAAGCAAGGAA 1237

seq_name: gb_rol:MUSTPHA

seq_documentation_block:
LOCUS MUSTPHA 1732 bp mRNA ROD 07-MAR-1995
DEFINITION Mouse tryptophan hydroxylase (Tph) mRNA, complete cds.
ACCESSION J04758
VERSION J04758.1 GI:202113
KEYWORDS tryptophan hydroxylase.
SOURCE Mouse (strain DBA; isolate P815) mastocytoma, cDNA to mRNA, clone
PGT3-113.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 1732)
AUTHORS Stoll,J., Kozak,C.A. and Goldman,D.
TITLE Characterization and chromosomal mapping of a cDNA encoding
tryptophan hydroxylase from a mouse mastocytoma cell line
JOURNAL Genomics 7 (1), 88-96 (1990)
MEDLINE 90243261
COMMENT Draft entry and computer-readable sequence for [Genomics (1990) In
press] kindly submitted
by J.Stoll, 26-JAN-1990.
FEATURES
source
1. 1732
/organism="Mus musculus"
/isolate="P815"
/strain="DBA"
/db_xref="taxon:10090"
/clone="PGT3-113"
/tissue_type="mastocytoma"
94. 1717
/gene="Tph"
94. 1437
/gene="Tph"
/EC_number="1.14.16.4"
/codon_start=1
/product="tryptophan hydroxylase"
/protein_id="AAA63401.1"
/db_xref="GI:202114"
/translation="MIEDNKKENKDHSSERGRVTLIFSLNEVGLIKVLKIFQEN
HVSLLHIESKRSKORNSERFIVDCDISREQLNDFPLKSHATVLSVSDPDLQTAKE
DVMETVPWPKKISDLDFCANRVLYSGDLADHPGKDNVYRRRKYFAELANNYKH
GDPIKTEFTETEEKTWGTVRELKLYPHACREYKLNPLITRYCGYREDNIPQLE
DVSNFLKRTGFSIRPVAGYLSPRDLAFLAFVHCTOYVRHSSDPLYTPEPDTCH
LIGHVPLLAEPFAQSFQETGLASLGADEAVOKLATCYFTVFEGFLCKQSGOLRV
```

AGLSSISLKHAKVFPDPKIAKOCILTSFQDVYFVSESEDAKEMREF  
AKTVKRPGLKYNPYTSQVQLRDKTSITAMNELRYDLVSDALARKTRWPSV"

polya\_signal 1712..1717

BASE COUNT 466 a 435 c 401 g 430 t

ORIGIN Chromosome 7.

alignment\_scores:  
Quality: 230.50 Length: 212  
Ratio: 1.889 Gaps: 6  
Percent Similarity: 57.547 Percent Identity: 31.132

alignment\_block:

US-09-438-185-1047 x MUSTPHA ..

Align seg 1/1 to: MUSTPHA from: 1 to: 1732

```
107 TrpTyrArgLeuLeuSerSerArgPheSerLeuTrpLysSerTyrCysPr 123
|||||
622 TGGGGGACATCTCCGAGAGCAACAACTTACCCGACCCACGCTG 671
123 oArgPheLeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheL 140
|||||
672 CAGG.....GAGTACCTCAGAACCTCCCTTGCCTCAAAATCT 712
140 euAspHis.....GlnAlaValIleLysPhe 148
|||||
713 GTGCTATATCGGAGAGCAACATCCGCACTGGAGGATGCTCCAACTTT 762
149 PheGluLeuGluThrHisPheSerTyrTyrProValSerGlyPheValAl 165
|||||
763 TTAAGAAGCACTGGGTTTCCATCGCTCTGCTGGCTGGTACCTCTC 812
165 aProHisGlnTyrLeuSerLeuLeuGlnAspArgTyrPheProIleAlas 182
|||||
813 ACCGAGAGATTTCTCTCGGGGTTAGCCTTTCGAGTCTTTCACCTGCAC 862
182 erValMetArgThrLeuAspLysAspAsnPheSerLeuThrProAspLeu 198
|||||
863 AGTATGTGAGACACACTTCAGATCCCTCTACACTCCAGACCCAGACACC 912
199 IleHisAspLeuLeuGlyHisValProThrPheLeuHisProSerPheSe 215
|||||
913 TCCATGATCTCTAGCCACGCTCTCTCTGCTGCAACCCAGTTTGC 962
215 rGluPhePheIleAsnMetGlyArgLeuPheThrLysValIleGluLysV 232
|||||
963 TCAATCTCCCAAGAAATGGC.....C 985
232 alGlnAlaLeuProSerLysLysGlnArgIleGlnThrLeuGlnSerAsn 248
986 TGGCTTCCTCTGGAGCTTCAGAGGAGACAGTTCAAAACTGGCAACG... 1032
249 LeuIleAlaIleValArgCysPheThrPheThrValGluSerGlyLeuI 265
|||||
1033 .....TGCTACTTTTTCAGCTGGAGTTTGGGCTGTG 1064
265 eGluAsnHisGluGlyArgLysAlaTyrGlyAlaValLeuIleSerSerP 282
|||||
1065 CAACAAGATGGACAGCTGAGAGCTTTTGGGCGCGCTGCTTCTTCTCCA 1114
282 rGlnGluLeuGlyHisAla.PheIleAspAsnValArgValLeuProLe 298
|||||
1115 TCAGTGACTCAACATGCACTTCTGAGATGCCAAAGTCAAGCCCTTT 1164
298 uGluLeuAspGlnIleLeuLeuLeuProPheAsn 309
|||||
1165 G.....ATCCAGATGCTGCTAAAC 1186
```

seq\_name: gb\_ba3:PSEPHY

seq\_documentation\_block:

LOCUS PSEPHY 2843 bp DNA BCT 02-MAY-1994  
DEFINITION Pseudomonas aeruginosa phenylalanine hydroxylase gene, complete cds.  
ACCESSION M88627  
VERSION M88627.1 GI:476740  
KEYWORDS phenylalanine hydroxylase.  
SOURCE Pseudomonas aeruginosa (strain PAO1) DNA.  
ORGANISM Pseudomonas aeruginosa  
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;  
Pseudomonas.  
REFERENCE 1 (bases 1 to 2843)  
AUTHORS Zhao,G.S., Xia,T., Song,J. and Roy,R.A.  
TITLE phenylalanine hydroxylase and 4 alpha-carbinolamine  
dehydratase/DCOH as part of a three-component gene cluster  
Proc. Natl. Acad. Sci. U.S.A. 91, 1366-1370 (1994)  
JOURNAL 94151331  
MEDLINE  
FEATURES  
Location/Qualifiers  
1..2843  
/organism="Pseudomonas aeruginosa"  
/strain="PAO1"  
/db\_xref="taxon:287"  
353..1141  
/standard\_name="ORF 1"  
/codon\_start=1  
/transl\_table=11  
/product="phenylalanine hydroxylase"  
/protein\_id="AA25936.1"  
/db\_xref="GI:476741"  
translation="MKTOYVARQPDNDNGTHYPTETHOVWNTLITROLKVIETRACO  
EYLDGELGHPERIPQIDRINVLQNTGHWARVPALIPQTEFELLSAQFPVA  
TFIRPELDLIGEDIFHEFGHCPLLTAPLAEFTHYGKLKASKERVFELRL  
YMTIEFGLVETDQGRRIIGGILSSKREIYVLSDEPHQAEFPLEANRTPYRIDIL  
QPIYFVLPLDKRLFLQAQEDIMALVHEANRLGLHAPLFFPKQA"  
1246..1602  
/codon\_start=1  
/transl\_table=11  
/product="phenylalanine hydroxylase"  
/protein\_id="AA25937.1"  
/db\_xref="GI:476742"  
translation="MTALTQAHCEACRADAPHVSDEELPVLLRQIPDWNIEVRDGM  
LEKYLKFNKHALAFTNAVEISEAGHHPGLLTENGKVTVTWWSHSTKGLHRNDFI  
MAATDEVAKTAEGRK"  
1599..2798  
/codon\_start=1  
/transl\_table=11  
/product="phenylalanine hydroxylase"  
/protein\_id="AA25938.1"  
/db\_xref="GI:476743"  
translation="MSHFAKVARVPGDPILGLDAYRNDPRADKLDLGVGVYKDAQGL  
TPILRSVKLAEORLVSEQTKSYVGSGDALFAALAEALGAASPLLLLEORADATQ  
PGGTGALRAGDFTAHCLPGIGLISDPTWPTHETFAAGLKVSHYPVPSADNRLDV  
EAMLAGLERIPQGVVLLHACCHNPTGFLSHDDMRVLDVVRRELPLIDFAYQGL  
GDLLEDAWAVRLFAGLEPELVLTSSCKNFGLYRDRVGALIVCAQNAEKLTDLRSOL  
AFIARNLWSTPBAHQAEVVAAILGDSKELGLMQEVEGMSRIASLRIGLVEALPHG  
LAERFHYGAGRGMSFTGLSQQVARNLDEHAVLYLVSSGRANVAGLHARRGLRLAQA  
IAQVCAD"

BASE COUNT 521 a 1016 c 846 g 460 t  
ORIGIN

alignment\_scores:  
Quality: 230.50 Length: 253  
Ratio: 1.568 Gaps: 7  
Percent Similarity: 58.103 Percent Identity: 24.506

alignment\_block:

US-09-438-185-1047 x PSEPHY ..

Align seg 1/1 to: PSEPHY from: 1 to: 2843

93 AspMetAlaValSerThrProPheAsnArgAsnLeuTrpTyrAr 109  
||| : : : : : ||| : : : : : |||

```

389 GACACGGTTTCATCCACTATCCGGAACCGGACCGACAGGTCTGGAATAC 438
109 gLeuLeuSerSerArgPheSerLeuTyrPlySerTyrCysProArgPheP 126
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
439 CTGTGATCCCGGCACTGAAGGTGATCGAAGCGCGCTGTGAGGAAT 488
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
126 heLeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheLeuAspHis 142
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
489 ACCTCGACGGCATCGACAGCTCGGCGGCGGCGGCGGCGGCGGCGGCGG 523
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
143 GlnAlaValIleLysPheGluLeuGlu.....Th 153
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
524 GACGGATCCCCAGCTCGACGATCAGAGATCAGAGGTCTCCAGCGCCACAC 573
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
153 rHisPheSerTyrTyrProValSerGlyPheValAlaProHisGlnTyrL 170
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
574 CGGCTGGCGCGTGGCACGGGTCTCGGCGGCTGATTCCTGTTCCAGACTTCT 623
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
170 euSerLeuLeuGlnAspArgTyrPheProIleAlaSerValMetArgThr 186
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
624 TCGAAGTGTGGCCAGCGACGAATTCCTCCGCTGCCACCTTCATCGGCACC 673
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
187 LeuAspLysAspAsnPheSerLeuThrProAspLeuIleHisAspLeuLe 203
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
674 CCGAAGAACTGGACTACCTACCGAGCGCGGACATCTTCCACGAGACTT 723
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
203 uGlyHisValProTyrPheLeuLeuHisProSerPheSerGluPheIleA 220
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
724 CGGCCACTGCCACCTGCTGACCAACCCCTGGCTCGCGGAGTTCCACCAT 773
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
220 sMetGlyArgLeuPheThrLysValIleGluLysValGlnAlaLeuPro 236
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
774 CTTACGGCAGCTCGGCTCAAGCGG..... 799
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
237 SerLysGlnArgIleGlnThrLeuGlnSerAsnLeuIleAlaIleA 253
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
800 AGCAAGGAGGACGC.....GTGTCTCTCGC 825
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
253 lArgCysPheThrPheThrValGluSerGlyLeuIleGluAsnHisGluG 270
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
826 CCCTCTGTTACTGATGACCATCGAGTTTCGGCCCTGGTGGAGACCGACAG 875
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
270 lArgLysAlaTyrGlyAlaValLeuIleSerSerProGlnGluLeuGly 286
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
876 GCAAGCGCATCTACGCGCGGCGCATCTCTCTCGCGGAGGAGACCGCTC 925
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
287 HisAlaPheIleAsp.....AsnValArgValLeuProLeuGluLe 300
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
926 TACAGCCTCTCGGAGCGCGCTGCACCGAGCTTCAATCCGCTGGAG... 973
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
300 uAspGlnIleIleArgLeuProPheAsnThrSerThrProGlnGluThrL 317
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
974 .....GGATGCGCGCGCCCTACCGCATCGACATCTCGCAACCGCTCT 1016
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
317 euPheSerIleArgHisPheAspGluLeuValGluLeuThrSerLysLe 333
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1017 ATTCTCTCTCGCCACCTCAAGCGCTGTTCACACTGTCGCCAGGAGAAC 1066
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
333 uGluTrp 335
||||| : : : : : : : : : : : : : : : : : : : : : : : : :
1067 ATCATGG 1073
||||| : : : : : : : : : : : : : : : : : : : : : : : : :

```

seq\_name: gb.pr10:HUMBTPH

```

seq_documentation_block: 1335 bp mRNA PRI 22-MAY-1995
LOCUS HUMBTPH Homo sapiens tryptophan hydroxylase (Tph) mRNA, complete cds.
DEFINITION L29306
VERSION L29306.1 GI:531192
KEYWORDS tryptophan hydroxylase.
SOURCE Homo sapiens brain stem cDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

REFERENCE
AUTHORS Tipper,J.P., Citron,B.A., Ribeiro,P. and Kaufman,S.
TITLE Cloning and expression of rabbit and human brain tryptophan
hydroxylase cDNA in Escherichia coli
JOURNAL Arch. Biochem. Biophys. 315 (2), 445-453 (1994)
MEDLINE 95077422
FEATURES
Location/Qualifiers
1..1335
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_line="E. coli DH5 (pJpHUMB)"
/tissue_type="brain stem"
1..1335
/gene="trph"
1..1335
/gene="trph"
/EC_number="1.14.16.4"
/Note="conflicts with Swissprot P17290 at amino acids 102,
151, 202, 207, and 390; protein phosphorylation at amino
acids 53, 58, 113, 260, 378, and 443; putative"
/codon_start=1
/product="tryptophan hydroxylase"
/protein_id="AAA67050.1"
/db_xref="GI:531193"
/translation="MIEDNKNKDHSLRGRATLIFSLKNEVGLIKALKIFQEKHVN
LLHIESRKSRRNSEFEIFVDCDTNREOLNDIFLLKSHNTNLSVTTPDNTMKKEGM
ESVPMFPKISLDLHCANRLMYGSELDPADHPGKDNVYRKRKYPADSAMYKYGDP
IPKVEETEEIKTWGTVFRELKLYPHACREYLKLNPLLSKYCGXQEDNIPQLEDIS
NFLKERTGESIRVAGYLSRDFLSGLAFRVPHCTQYVRHSSDPFYTPEDPTCHGLL
HYELLAPSPAFOSRIGLASGSEAAVKLATCYFTVERGLCKDGLRVFGAGL
LSSISELKHVLSHAKYRPFDPKIKTCQEGCLITTFQDVYEVSESEFENAEKREFTKT
IKRFPGVKNPITRSIQILDKAKSITKNWELRDLDDVSDALGKVSRRKPSI"
BASE COUNT 404 a 283 c 283 g 365 t
ORIGIN

```

```

alignment_scores:
Quality: 229.50 Length: 230
Ratio: 1.752 Gaps: 4
Percent similarity: 56.957 Percent identity: 27.391

```

alignment\_block:

US-09-438-185-1047 x HUMBTPH

Align seg 1/1 to: HUMBTPH from: 1 to: 1335

```

107 TriPtyrArgLeuLeuSerSerArgPheSerLeuTyrPlySerTyrCysPr 123
||| : : : : : : : : : : : : : : : : : : : : : : : : :
520 TGGGGAACCGTATTCGGGAGCTCAACAACTCTATCCGACCATGCTTG 569
||| : : : : : : : : : : : : : : : : : : : : : : : : :
123 oAtgPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheL 140
||| : : : : : : : : : : : : : : : : : : : : : : : : :
570 CAGA.....GAGTATCTCAAAATTTACCTCTGCTTTCCAAGTATT 610
||| : : : : : : : : : : : : : : : : : : : : : : : : :
140 euAspHisGlnAla.....ValIleLysPhe 148
||| : : : : : : : : : : : : : : : : : : : : : : : : :
611 GTGGATATCAGGAAGACAATATCCACAGCTGGAGATATTTCAAACTTT 560
||| : : : : : : : : : : : : : : : : : : : : : : : : :
149 PheGluLeuGluThrHisPheSerTyrTyrProValSerGlyPheValAl 165
||| : : : : : : : : : : : : : : : : : : : : : : : : :
661 TTTAAAGAGCGCACAGGTTTTTCCATTCGCTCTGCTGCTGTTACTTATC 710
||| : : : : : : : : : : : : : : : : : : : : : : : : :
165 aproHisGlnTyrLeuSerLeuGlnAspArgTyrPheProIleAlaAs 182
||| : : : : : : : : : : : : : : : : : : : : : : : : :
711 ACCAAGAGATTTTATCATAGGTTTAGCCCTTTTCGAGTTTTCACCTGC 760
||| : : : : : : : : : : : : : : : : : : : : : : : : :
182 erValMetArgThrLeuAspLysAspAsnPheSerLeuThrProAspLeu 198
||| : : : : : : : : : : : : : : : : : : : : : : : : :
761 AATATGTGACACACAGTTCACACCCCTTCTATACCCACGAGCGGATACC 810
||| : : : : : : : : : : : : : : : : : : : : : : : : :
199 IleHisAspLeuLeuGlyHisValProTrpLeuLeuHisProSerPheSe 215
||| : : : : : : : : : : : : : : : : : : : : : : : : :

```



811 TGCCATGAACCTCTTAGGTACACCTGCCCTTCCCTTTGGCTGAGCAAGTTTTC 860

215 rGlupPhePheIleAsnMetGlyArgLeuPheThrLysValIleGluLysV 232

861 TCAGTTCTCCCAAGAAATTGGC.....C 883

232 alGlnAlaLeuProSerLysLysGlnArgIleGlnThrLysLysSerAsn 248

884 TGGCTTCCCTTGGAGCTTCAGAGAGCTGTTCAAAACCTGGCAACG... 930

249 LeuIleAlaIleValArgCysPheThrPheThrValGluSerGlyLeuI 265

931 .....TGCCTACTTTTCACTGTGGAGTTTGGTCTATG 962

265 eGluAsnHisGluGlyArgLysAlaTyrGlyAlaValLeuIleSerSerp 282

963 TAAACAAGACGACAGTACGAGTCTTCGGCGCTGCTTACTTCTTCTTA 1012

282 roGlnGluLeuGlyHisAlaPheIleAsnValArgValLeuProLeu 298

1013 TCAGTGAACCTCAACATGTCTTCTGGACATGCCAAAGTAAGCCCTTT 1062

299 GluLeuAspGlnIleIleArgLeuProPheAsnThrSerThrProGlnI 315

1063 GATCCCAAGATTACCTGCAACAAAGAAATGCCTCAATCAACACTTTTCAGGA 1112

315 uThrLeuPheSerIleArgHisPheAspGluLeuValGlu 328

1113 TGTCTACTTTGTATCTGAAGTTTTCGAAGATGCAAAAGGAG 1152

seq\_name: gb\_om:RABTPH

seq\_documentation\_block: 1412 bp mRNA MAM 22-MAY-1995

LOCUS RABTPH Oryctolagus cuniculus cuniculus (Tph) mRNA, complete cds.

DEFINITION Oryctolagus cuniculus hind and midbrain cDNA to mRNA.

ACCESSION L29305.1 GI:531213

VERSION L29305.1

KEYWORDS Oryctolagus cuniculus

SOURCE Oryctolagus cuniculus

ORGANISM Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.

REFERENCE 1 (bases 1 to 1412)

AUTHORS Tipper, J.P., Citron, B.A., Ribeiro, P. and Kaufman, S.

TITLE Cloning and expression of rabbit and human brain tryptophan hydroxylase cDNA in Escherichia coli

JOURNAL Arch. Biochem. Biophys. 315 (2), 445-453 (1994)

MEDLINE 95077422

SOURCE Location/Qualifiers

1. 1412

/organism="Oryctolagus cuniculus"

/db\_xref="taxon:9986"

/cell\_line="E. coli DH5 (clone p1pRAB)"

/tissue\_type="hind and midbrain"

1. 1335

/gene="Tph"

1. 1335

/gene="Tph"

/EC\_number="1.14.16.4"

/note="conflicts with Swissprot P17752 at amino acids 19, 68, 90, 97, 100, 104, 151, 154, 157, 179, 207, 217, 344, 414, 419, 425 and 436; protein phosphorylation at amino acids 53, 58, 113, 260, 378 and 443; putative"

/codon\_start=1

/product="tryptophan hydroxylase"

/protein\_id="AAA67051.1"

/db\_xref="GI:531214"

/translation="MTEDNKNKDSHSLGRATLIFSLKNEVGGLIKALKIFQKHVN LLHTEKSKRRNSFEIFVDCDNRQLNDFLLKSHNTNLSVTTPDNTMKEGM ESWPFPKTSIDLDHCANRLVMGSELDADPGFKDNYRKRKYFADSAMSYRGDP IPKVTFEETIKWTGTVRELNKLKYPHACREYLKRLSKYCGYQEDNIPQLEDIS NFLKRTGFSIRPVAGYLSRDELGLAFAVFHCTQYVRHSSDPFTTPEPTCHLLG

gene

CDS

BASE COUNT 416 a 306 c 305 g 385 t

ORIGIN

alignment\_scores: Quality: 229.50 Length: 230

Ratio: 1.752 Gaps: 4

Percent Similarity: 56.957 Percent Identity: 27.391

alignment\_block:

US-09-438-185-1047 x RABTPH ..

Align seg 1/1 to: RABTPH from: 1 to: 1412

107 TriPtyrArgLeuLeuSerSerArgPheSerLeuTriPlyssSerTyrCysPr 123

520 TGGGGAACCGTATTCCGGGAGCTCAACAACTCTATCCGACCATGCTTG 569

123 oAtqPhePheLeuAspTyrLeuGluAlaPheGlyLeuLeuSerAspPheL 140

570 CAGA.....GAGTATCTCAAAATTTACCTCTGCTTTCCAAGTATT 610

140 euAspHisGlnAla.....ValIleLysPhe 148

611 GTGATATACGAGGAACAATATCCACAGCTGGAGATATTTCAACTTT 660

149 PheGluLeuGluThrHisPheSerTyrTyrProValSerGlyPheValal 165

661 TTAAGAAGCGCACAGGTTTTTCCATCTGCTCTGTGGCTGTACTATTC 710

165 aProHisGlnTyrLeuSerLeuGlnAspArgTyrPheProIleAlas 182

711 ACCAAGAGATTTCTTATAGGTTTAGCCCTTTTCGAGTTTTTCACTGCACT 760

182 erValMetArgThrLeuAspLysAspAsnPheSerLeuThrProAspLeu 198

761 AATATGTGACACACAGTTCACACCCCTTCTATACCCAGAGCCGGATACC 810

199 IleHisAspLeuGlyHisValProTrpLeuLeuHisProSerPheSe 215

811 TCCCATGAACCTTTAGGTACAGTTCCTTTTGGCTGAGCAAGTTTTC 860

215 rClupPhePheIleAsnMetGlyArgLeuPheThrLysValIleGluLysV 232

861 TCAGTTCTCCCAAGAAATTGGC.....C 883

232 alGlnAlaLeuProSerLysLysGlnArgIleGlnThrLysLysSerAsn 248

884 TGGCTTCCCTTGGAGCTTCAGAGAGCTGTTCAAAACCTGGCAACG... 930

249 LeuIleAlaIleValArgCysPheThrPheThrValGluSerGlyLeuI 265

931 .....TGCCTACTTTTCACTGTGGAGTTTGGTCTATG 962

265 eGluAsnHisGluGlyArgLysAlaTyrGlyAlaValLeuIleSerSerp 282

963 TAAACAAGACGACAGTACGAGTCTTCGGCGCTGCTTACTTCTTCTTA 1012

282 roGlnGluLeuGlyHisAlaPheIleAsnValArgValLeuProLeu 298

1013 TCAGTGAACCTCAACATGTCTTCTGGACATGCCAAAGTAAGCCCTTT 1062

299 GluLeuAspGlnIleIleArgLeuProPheAsnThrSerThrProGlnI 315

1063 GATCCCAAGATTACCTGCAACAAAGAAATGCCTCAATCAACACTTTTCAGGA 1112

315 uThrLeuPheSerIleArgHisPheAspGluLeuValGlu 328

1113 TGTCTACTTTGTATCTGAAGTTTTCGAAGATGCAAAAGGAG 1152



[illegible]

BASE COUNT	368 a	531 c	472 a	386 t
/gene= IN				

	368 a	531 c	472 g	386 t
BASE COUNT				
ORIGIN				

alignment_scores:		
Quality:	225.50	Length: 240
Ratio:	1.658	Gaps: 5
Percent Similarity:	56.667	Percent Identity: 27.500

alignment\_block:  
US-09-438-185-1047 x MUSTHRA ..  
..... 1/1 to: MUSTHRA from: 1 to: 1757

to: 1757  
from: 1

107 TriptyrArgLeuLeuSerSeraArgPheSerLeuTrpLysSerTyrCysPr 123  
 :  
 702 TGG AAGGAGGTATACCGCACGCCTGAAGGCCCTATTACTACCATACCCATCCGCG 751  
 :  
 123 oAgrpHePheLeuAspTyTrLeuGLuaIahpeGLyLeuLeuSerAspPheL 140  
 :  
 752 CCGG ..... GAACAACCTGGAGCTTTTCAGCTCTTGCAACGGTA 792  
 :  
 140 euASpHis ..... GlnAlavalIIeLYSPhe 148  
 :  
 793 GTGGCTACCGAGGAGGACAGCATTCACAGCTGGAGGATGTGTCTCACTTC 842  
 :  
 149 PheGLinLeuLnThHispheSeTyTrTyProValserGLYpheValal 165  
 :  
 843 TTGAAGAACCGAGTGGCTTCACAGCTGCCAGCGTGGCGGCTACTGTGTC 892  
 :  
 165 apRoHisLnTyLeuSerLeuLeuGLInasParqTyRheProIleAalas 182  
 :  
 893 TGCCCGTGATTTCTGGCCAGTCTGGCCTTCCTCGTGTGTTTCAGTGCACAC 942  
 :  
 182 erValKetArgThrLeuasPLYSASPasnPheserLeuthTroaspLeu 198  
 :  
 943 AGTACATCCGTGTCATGCTCCTTCACCTATGCACACTCCCGAGCCAGACTGC 992  
 :  
 199 IleHisapLeuLeuGLyHisVaIProTpLeuLeuHisProSerPheSe 215  
 :  
 993 TGCCACAGAGCTGCTGGGACACGTACCNCATGTTGGCTGACCGCAGCATTGC 104  
 :  
 215 rGLUphePheIleasnMeCLyArgLeupheThrySLysValeIGluySV 232  
 :  
 1043 CCAGTTCCTCCCAGACATTTGA.....C 106  
 :  
 232 alclnalaLeuproSerLysLysGLnarGLieGLInthrLeuGLnsarSn 248  
 :  
 1066 TTGCATCTCTGGGGCTTCAGATGAAGAANTTAAANCTCTCCACG... 111  
 :  
 249 LeuilleAlaIleValArgCysPheTrpPheThValcILserGLyleuIl 265  
 :  
 1113 .....GTGTACTGGTTCACCTGTGAGATTTGGGCTGTG 114  
 :  
 265 eGLUASNHisGLUGlyArgLySalatYrgLYalavaLeulleSerSerP 287  
 :  
 1145 TAAACAGAAATGGGAGCTGGAAGCTTACGGTGACAGGCTGCTGTCTTCT 116  
 :  
 282 roGLInGLuLeuGLyHIISalapheIleASPsnValArGVallLeuProLeu 298  
 :  
 1195 ATVGAGAGAGCTCTCGCATCTCCCTGTCAGAGAGCCGAGGTCCGGGCCCTT 120  
 :  
 299 GLutLeuASPGLInIleIArGLeuProPheAsnthrrSerThProGLnGl 316  
 :

Mon Jul 2 08:46:42 2001

us-09-438-185-1047.rge

---

Page 20



GenCore version 4.5  
Copyright (C) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2001, 13:48:36 ; Search time 58.56 Seconds  
(without alignments)  
816.475 Million cell updates/sec

Title: US-09-438-185-1047

Perfect score: 1889

Sequence: 1 VHCERTLDPKYILKALKL.....ESIFLYNQKYLGSFEVLQ 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL16:\*\*

- 1: sp\_archaea:\*\*
- 2: sp\_bacteria:\*\*
- 3: sp\_fungi:\*\*
- 4: sp\_human:\*\*
- 5: sp\_invertebrate:\*\*
- 6: sp\_mammal:\*\*
- 7: sp\_mhc:\*\*
- 8: sp\_organelle:\*\*
- 9: sp\_phase:\*\*
- 10: sp\_plant:\*\*
- 11: sp\_rodent:\*\*
- 12: sp\_unclassified:\*\*
- 13: sp\_vertebrate:\*\*
- 14: sp\_virus:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1886	99.8	362	2	Q9Z6L3	Q9Z6L3 chlamydia p
2	256.5	13.6	289	2	Q9KLB8	Q9KLB8 vibrio chol
3	237.5	12.6	491	13	Q9PU40	Q9PU40 gallus gall
4	224.5	11.9	497	5	Q96370	Q96370 schistosoma
5	219	11.6	579	5	Q24000	Q24000 drosophila
6	206.5	10.9	532	5	Q9XZD1	Q9XZD1 caenorhabdi
7	206.5	10.9	575	5	Q23438	Q23438 caenorhabdi
8	196.5	10.4	297	2	Q9XC88	Q9XC88 chromobacte
9	181.5	9.6	457	5	Q9XY05	Q9XY05 caenorhabdi
10	178.5	9.4	450	5	Q96947	Q96947 geodia cydo
11	178.5	9.4	555	5	Q9W0K2	Q9W0K2 drosophila
12	177	9.4	495	2	Q9W0K2	Q9W0K2 myxococcus
13	157	8.3	438	5	Q17498	Q17498 branchiosto
14	109.5	5.8	1253	10	Q9SCZ3	Q9SCZ3 arabidopsis
15	106.5	5.6	506	8	Q9GF63	Q9GF63 arabis alpi
16	106	5.6	129	13	Q42428	Q42428 lates calca
17	106	5.6	504	8	Q9GF55	Q9GF55 cardamine a
18	106	5.6	504	8	Q9GF31	Q9GF31 cardamine p
19	106	5.6	504	8	Q9GF30	Q9GF30 cardamine r

20	103.5	5.5	1782	5	Q9VPI9	Q9VPI9 drosophila
21	103	5.5	504	8	Q9GF41	Q9GF41 aubrieta de
22	103	5.5	513	8	Q9TIB8	Q9TIB8 bouteloua c
23	102.5	5.4	1154	5	Q9VM23	Q9VM23 drosophila
24	102	5.4	506	8	Q9GF64	Q9GF64 arabis alpi
25	101.5	5.4	519	8	Q9MV23	Q9MV23 oryza meyer
26	101	5.3	95	2	Q9R634	Q9R634 chromobacte
27	101	5.3	2042	5	Q9WIC5	Q9WIC5 drosophila
28	100.5	5.3	102	11	P97517	P97517 phodopus su
29	100.5	5.3	2212	14	Q9DQD1	Q9DQD1 ebola virus
30	99.5	5.3	513	8	Q9TI97	Q9TI97 sporobolus
31	99	5.2	924	11	O54921	O54921 rattus norv
32	98.5	5.2	513	8	Q9TIC0	Q9TIC0 zeugites pi
33	98.5	5.2	584	2	Q9PL28	Q9PL28 campylobact
34	98	5.2	1887	5	Q9GZ66	Q9GZ66 c contains
35	97.5	5.2	239	5	Q9NJQ3	Q9NJQ3 lymnaea sta
36	97.5	5.2	504	8	Q9GF36	Q9GF36 lepidium ca
37	97.5	5.2	519	8	Q9MV31	Q9MV31 oryza alta
38	97.5	5.2	519	8	Q9MV30	Q9MV30 oryza latif
39	97.5	5.2	519	8	Q9MV29	Q9MV29 oryza grand
40	97	5.1	347	4	Q13786	Q13786 homo sapien
41	97	5.1	504	8	Q9GF53	Q9GF53 arabis proc
42	97	5.1	504	8	Q9GF28	Q9GF28 arabidopsis
43	97	5.1	519	8	Q9MV25	Q9MV25 oryza ridle
44	97	5.1	519	8	Q9MV24	Q9MV24 oryza longi
45	96.5	5.1	363	8	Q9GHI8	Q9GHI8 caryodapimo

#### ALIGNMENTS

RESULT 1

ID	Q9Z6L3	PRELIMINARY;	PRT;	362 AA.
AC	Q9Z6L3;			
DT	01-MAY-1999 (TEMBLrel. 10, Created)			
DT	01-MAY-1999 (TEMBLrel. 10, Last sequence update)			
DT	01-MAR-2001 (TEMBLrel. 16, Last annotation update)			
DE	AROMATIC AMINO ACID HYDROXYLASE.			
GN	CPN1046 OR CPJ1045 OR CP0806			
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).			
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia			
OX	NCBI_TaxID=83558;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CN1029;			
RX	MEDLINE=99206606; PubMed=10192388;			
RA	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,			
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.;			
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis."			
RL	Nat. Genet. 21:385-389(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=J138;			
RX	MEDLINE=20330349; PubMed=10871362;			
RA	Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,			
RA	Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;			
RT	"Comparison of whole genome sequences of Chlamydia pneumoniae J138			
RT	from Japan and CN1029 from USA."			
RL	Nucleic Acids Res. 28:2311-2314(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=AR39;			
RX	MEDLINE=20150255; PubMed=10684935;			
RA	White T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,			
RA	White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,			
RA	Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,			
RA	Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J.,			
RA	McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;			
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia			
RT	pneumoniae AR39."			
RL	Nucleic Acids Res. 28:1397-1406(2000).			
RN	EMBL; AE001685; AAD19183.1; -			

```
DR EMBL; AF002548; BAA9253.1; -
DR TIGR; AE002240; AAF73705.1; -
DR InterPro; IPR001273; -
DR Pfam; PF00351; bioterin_H; 1.
DR PRINTS; PR00372; FYWHYDRXLASE.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
SQ SEQUENCE 362 AA; 42513 MW; 01B89B4B4FE593B CRC64;

Query Match          99.8%; Score 1886; DB 2; Length 362;
Best Local Similarity 99.7%; Pred. No. 4.6e-150;
Matches 361; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VHCERTLDPKYILKIALKROSLFQNSQSLQRAYSTPYSYRITLLOKREKQALA 60
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MYHCERTLDPKYILKIALKROSLFQNSQSLQRAYSTPYSYRITLLOKREKQALA 60
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

( 61 RHKCSILEFFKNLLFVLLSLSKNQREGSCDMVSTPFFNRNLWRLSSRFLWS 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 RHKCSILEFFKNLLFVLLSLSKNQREGSCDMVSTPFFNRNLWRLSSRFLWS 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 121 YCPRFLDYLEAFGLSLDFLDHQAVIKFPELETHFSYYPVSGFVAPHOYLSLQDRYFPI 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 YCPRFLDYLEAFGLSLDFLDHQAVIKFPELETHFSYYPVSGFVAPHOYLSLQDRYFPI 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 181 ASVMRTLDKDNFSLTPDLIHLGHVFWLLHPSFSFEFFINMGRLETKVIEKVOALPSKKQ 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 ASVMRTLDKDNFSLTPDLIHLGHVFWLLHPSFSFEFFINMGRLETKVIEKVOALPSKKQ 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 241 RIQTQSLNLIATVRCFWFTVSGLIENHEGRKAYGAVLISSPQELGHAFIDNVRVLPLEL 300
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 RIQTQSLNLIATVRCFWFTVSGLIENHEGRKAYGAVLISSPQELGHAFIDNVRVLPLEL 300
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 301 DQIIRLPNTSPQETLSIRHFDVELSKLEWMLDGLLESTPLYNOEKLGGFEVL 360
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 DQIIRLPNTSPQETLSIRHFDVELSKLEWMLDGLLESTPLYNOEKLGGFEVL 360
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 361 CQ 362
   :|||:
Db 361 CQ 362
   :|||:
```

```
DR Pfam; PF00351; bioterin_H; 1.
DR PRINTS; PR00372; FYWHYDRXLASE.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
SQ SEQUENCE 289 AA; 33445 MW; 2D68B31C6E31D521 CRC64;

Query Match          13.6%; Score 256.5; DB 2; Length 289;
Best Local Similarity 25.6%; Pred. No. 1e-13;
Matches 65; Conservative 56; Mismatches 108; Indels 25; Gaps 6;

QY 106 LWYLLSSRFLSKSYCPREFLDYLEAFGLSLDFLDHQAVI-KFELETHFSYYPVSGFV 164
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 49 VHEHLTRQEVVKTACQAYLDGLNMLNLPDRLPQLPNEINRLQRETGWQVEPVPALI 108
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 165 APOHLYSLQDRVFPTASVMRTLDKDNFSLTPDLIHLGHVFWLLHPSFSFEFFINMGR 224
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 109 SFDFPALLADKKFPVATFLRRREEDYLOEDFEEFVYGHCMALTPDPAATHYVQGL 168
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 225 FTKVIEKVOALPSKKQRIQTQSLNLIATVRCFWFTVSGLIENHEGRKAYGAVLISSPQE 284
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 169 GAKATPKERSY-----LARYWFTVEFGLVQEGQTKIYGGGLSSPGE 212
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 285 LGHAFIDNV-RVLPLELDQIIRLPNTSPQETLSIRHFDVELTSK-----LEWM-LD 338
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 213 TLYASESTIPKREFPDIMQVLRTPYRIDIMQPIYVLPDLSQLYQLSQRDVMAVWQAMQ 272
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

QY 339 QGLLESIPLYNOEK 352
   :|||:|||||:
Db 273 DGLLP--PLFQPK 284
   :|||:|||||:

RESULT 3
QSPU40 PRELIMINARY; PRT; 491 AA.
AC Q9PU40;
DT 01-MAY-2000 (T-EMBLrel. 13, Created)
DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (T-EMBLrel. 15, Last annotation update)
DE TYROSINE HYDROXYLASE.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Ernsberger U.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96076133; PubMed=7577670;
RA Ernsberger U., Patzke H., Tissier-Seta J.P., Reh T., Goridis C.,
RA Rohrer H.;
RT "The expression of tyrosine hydroxylase and the transcription factors
RT cPhox-2 and Cash-1: evidence for distinct inductive steps in the
RT differentiation of chick sympathetic precursor cells.";
RL Mech. Dev. 52:125-136(1995).
DR EMBL; AJ251387; CAB62388.1; -.
DR HSSP; P04177; ITOH.
DR InterPro; IPR001273; -.
DR Pfam; PF00351; bioterin_H; 1.
DR PRINTS; PR00372; FYWHYDRXLASE.
DR PROSITE; PS00367; BIOTERIN_HYDROXYL; 1.
SQ SEQUENCE 491 AA; 56016 MW; 6322F2D58E746930 CRC64;

Query Match          12.6%; Score 237.5; DB 13; Length 491;
Best Local Similarity 29.1%; Pred. No. 7.8e-12;
Matches 66; Conservative 36; Mismatches 98; Indels 27; Gaps 4;

QY 107 WYRLSSRFLSKSYCPREFLDYLEAFGLSLDFLDH-----QAVIKFPELETHFSY 158
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 226 WKEVYSTLSLYPTHACK---EYLEAFNLLEKFCYNNENIPQLEEVSRFLKERTGQLR 282
   :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
```

```
QY 159 PVSGFVAPHOYLSLLQDRYFIAPSVMTLKDKNFSLTPDLIHLLGHVPLWLLHSPSEFF 218
DB 283 PVAGLSARDFLASLAFRVCTQYIRHASPMSPEPCCHLGHVPLMLADTFAQFS 342
QY 219 INNGRLFTVIEKVALPSKQRTOTLOSNLIAIVRCFWFTVESGLIENHGRKAYGAVL 278
DB 343 ODIG-----LASLGATEDEEKLATL-----YWFTEFGLCRQNGIVKAYGAGL 386
QY 279 ISSPQELGHAFIDNVRLPLEDLQIIRLPTNTPTQETLFSIRHDE 325
DB 387 LSSYGLIHLSLDEPEVDFDPDAAAVQYQDQYQPVYFVSEFSD 433

RESULT 4
ID O96370 PRELIMINARY: PRT; 497 AA.
AC O96370;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
TP TRYPTOPHAN HYDROXYLASE (EC 1.14.16.4).
OS Schistosoma mansoni (Blood fluke).
OC Eukaryota; Metazoa; Platyhelminthes; Rhabditophora; Neodermata;
OC Trematoda; Digenea; Strigoida; Schistosomatidae; Schistosomatidae;
OX NCBI_TaxID=6183;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PUERTO RICAN;
RA Hamdan F.F., Ribeiro P.;
RT "Molecular cloning and characterization of tryptophan hydroxylase from
RT Schistosoma mansoni."
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF031034; AAD01923.1; -.
DR HSSP; P04177; ITOH.
DR InterPro; IPR001273; -.
DR Pfam; PF00351; bioperlin_H; 1.
DR PRINTS; PR00372; FYWHYDRXLASE.
DR PROSITE; PS00367; BIOPERLIN_HYDROXYL; 1.
KW Oxidoreductase.
SQ SEQUENCE 497 AA; 57598 MW; F8964E4B2C361D CRC64;

Query Match 11.98; Score 224.5; DB 5; Length 497;
Best Local Similarity 28.5%; Pred. NO. 9.7e-11;
Matches 68; Conservative 34; Mismatches 108; Indels 29; Gaps 5;

96 VVSTPFNNMRYLLSRFLSKW-SYCPREFLDYLEAFGLSDFLDH-----QAVI 146
DB 201 IVEYTEIKTGWRIYRELTRYKTSACHEF---QKNGLQLQKAGYNEFDLPQLQWS 256
QY 147 KFELETHFSYYPVSGFVAPHOYLSLLQDRYFIAPSVMTLKDKNFSLTPDLIHLLGHV 206
DB 257 DFLKARTGCLRPVAGYLSARDLGLAFRVFYCTQYIRHQADPEYTPPCCHLGHV 316
QY 207 PVLHPSFSEFFINMGRFLTIVKIEKVALPSKQRTOTLOSNLIAIVRCFWFTVESGLIE 266
DB 317 PMLADPKFARFSEIG-----LASLQTSDEEIKLAT-----CYFFTIEFGLCR 360
QY 267 NHEGRKAYGAVLISPPQELGHAFIDNVRLPLEDLQIIRLPTNTPTQETLFSIRHDE 325
DB 361 QDNOLKAYCAGLSSVAELQHALSKAVIRFPMPKVINCECLVTFQNGYFEFSSFD 419

RESULT 5
ID Q24000 PRELIMINARY: PRT; 579 AA.
AC Q24000;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
```

```
DE PLE PROTEIN
GN PLE OR CG10118.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Prorygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anandides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Adayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhattacharya P., Brottier P.,
RA Borkova D., Botchan M.R., Bouch J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush C., Kraft C., Kravitz S., Kulp D., Kulp D.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Teh K.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON S;
RA MEDLINE=95014502; PubMed=7929381;
RX Birman S., Morgan B., Anzivino M., Hirsh J.;
RT "A novel and major isoform of tyrosine hydroxylase in Drosophila is
RT generated by alternative RNA processing."
RL J. Biol. Chem. 269:36559-36567(1994).
RP [3]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON S;
RA MEDLINE=90166583; PubMed=2483109;
RX Neckameyer W.S., Quinn W.G.;
RT "Isolation and characterization of the gene for Drosophila tyrosine
RT hydroxylase."
RL Neuron 2:1167-1175(1989).
DR EMBL; AE003561; AAF50648.1; -.
DR EMBL; U14395; AAG2876.1; -.
DR HSSP; P04177; ITOH.
DR FlyBase; FBgn0005626; ple.
DR InterPro; IPR001273; -.
DR Pfam; PF00351; bioperlin_H; 1.
DR PRINTS; PR00372; FYWHYDRXLASE.
DR PROSITE; PS00367; BIOPERLIN_HYDROXYL; 1.
```





Db 411 QFSQETG-----LASIGASEEDLKKLATL-----YFSTIERGLSSDDAASPVK 454  
 QY 266 ---ENHEGRKAYGAVLISSPOELGHAFIDNVRVLPLELDQIIRLPFNFTSTPQETLSIRH 322  
 Db 455 ENSGNHERKVVYAGLLSSAGELOHAVESATIIIRPDPRVVEOELTITTFQSAFYFTNR 514  
 QY 323 FDELVELTSKLEWMLDQGLLESLYNOEKYLSGFEVL 360  
 Db 515 FEEAQO---KLRFMTNNMKRPFIVRYN--PYTESVEVL 547

## RESULT 8

O9XC88  
 ID O9XC88 PRELIMINARY; PRT; 297 AA.  
 AC O9XC88;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)  
 DE PHENYLALANINE HYDROXYLASE.  
 OS Chromobacterium violaceum.  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae;  
 Chromobacterium.  
 NCBI\_TaxID=536;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-AFCC 12540;  
 RA Volner A., Nersisyan A.M., Abu-Omar M.M.;  
 RT "Expression, isolation, and metal-dependent catalysis of phenylalanine  
 hydroxylase from Chromobacterium violaceum."  
 RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF146711; AAD37774.1; --  
 DR HSSP: P00439; 4PAH.  
 DR InterPro: IPR001273; --  
 DR Pfam: PF00351; biopterin\_H; 2.  
 DR PRINTS: PR00372; FWHYDRXLASE.  
 DR PROSITE: PS00367; BIOPTERIN\_HYDROXYL; 1.  
 SQ SEQUENCE 297 AA; 33594 MW; AAA67B6097171FB0 CRC64;

Query Match 10.4%; Score 196.5; DB 2; Length 297;  
 Best Local Similarity 27.4%; Pred. No. 1.1e-08;  
 Matches 63; Conservative 40; Mismatches 94; Indels 33; Gaps 6;

QY 126 FIDYLEAFGLSDFL-DHQAVIKFELETHTSPYSGVAPHOYLSLLQDRYFFIASVM 184  
 Db 63 FLEGLELEVADADRVDFNKLMEKMAATGWKIVAPGLIPDDVFFHEHLANRRFFVTWML 122  
 QY 185 RTLDKDNFSLTPDLIHDLLGHVPLLPSPSEFFINMRLFTKVIKVOALPSKKQRIQT 244  
 Db 123 REPHQLDYLOEPDVFHDLFGHVPLLPINVFADYLEAYGKGVKR-KALGALP----- 173  
 QY 245 LQSNLIAIVRCVFTVESGLIENHEGRKAYGAVLISSPQE---LGHAFIDNVRVLPLEL 300  
 Db 174 -----MLARLYNTVTFEGLINTPAGMRYGAGILSSKSESIYCLDSASPNRV---GFDL 224  
 QY 301 DQIIRLPNTSTPQETLSIRHFDVELTSKLEWMLDQGLLESIPLYNQ 350  
 Db 225 MEIMNTRYRIDTQKTYFVIDSKQLFDATAP-----DFAPLYLQ 264

## RESULT 9

O9AYQ5  
 ID O9AYQ5 PRELIMINARY; PRT; 457 AA.  
 AC O9AYQ5;  
 DT 01-NOV-1999 (Tremblrel. 12, Created)  
 DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
 DT 01-NAR-2001 (Tremblrel. 16, Last annotation update)  
 DE PHENYLALANINE HYDROXYLASE (EC 1.14.16.1).  
 GN K0888.4.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 NCBI\_TaxID=6239;

RN SEQUENCE FROM N.A.  
 RP STRAIN-BRISTOL N2;  
 RC "Joel C.M., Davidson B., McKerrow J.;  
 RA "A phenylalanine hydroxylase gene from the nematode Caenorhabditis  
 elegans is expressed in the hypodermis."  
 RT J. Neurogenet. 0:0-0(1999).  
 KL EMBL: AF119388; AAD51643.1; --  
 DR HSSP: P00439; 4PAH.  
 DR InterPro: IPR001273; --  
 DR InterPro: IPR001273; --  
 DR Pfam: PF00351; biopterin\_H; 1.  
 DR PRINTS: PR00372; FWHYDRXLASE.  
 DR PROSITE: PS00367; BIOPTERIN\_HYDROXYL; 1.  
 DR PROSITE: PS00192; CYTOCHROME\_B\_HEME; UNKNOWN\_1.  
 KW Oxidoreductase.  
 SQ SEQUENCE 457 AA; 52189 MW; 7A573B884B9EF6FC CRC64;

Query Match 9.6%; Score 181.5; DB 5; Length 457;  
 Best Local Similarity 21.8%; Pred. No. 3.5e-07;  
 Matches 83; Conservative 66; Mismatches 137; Indels 95; Gaps 15;

QY 14 LKIALKRLSLSLFONSSLOLAY----STPVSYRILLOKENKEQALARHKCISILE 69  
 Db 81 LKTIIVRLKELLISI---SNKLRFRLEKGTGTPK-----KONKDSYVPFPKINDIDQ 130  
 QY 70 FFKNLL-----FVHLLSKNQREGSTOMA-----VVSTFFFNRL--WYRL 110  
 Db 131 FANRILSYGAELDADHPGKDMTYRERKFFADIAENFKHGDKIPIITYTDEETATWTV 190  
 QY 111 LSSRFLW-KSYCPREFLDYLEAFGLLSD-----FLDHOAVIKFELETHTSPYSPVS 161  
 Db 191 YNLTVMYKNAQOE--NYI--FPLQNCNGFGPDRIPQLODVSDFLKDCGTGTYIRPA 246  
 QY 162 GFVAPHQYLSLLQDRYFFIASVMRTLDKDNFSLTPDLIHDLLGHVPLLPSPSEFFINM 221  
 Db 247 GLLSRDFLAGWAFRVFHSQYIRHSAKRYTPPEPDICHELLGHVPLFADYVFAQFSOI 306  
 QY 222 GRLF----TKVTEKQVLPSPKKQRIQTQLSNLIAIVRCVFTVESGLIENHEGRKAYGAV 277  
 Db 307 GLASLAPDDVIEKLATL-----YWFTEFGICQDQGEKKAYGAG 346  
 QY 278 LISSPOELGHAFIDNVRVLPLE--LDQIIRLPNTSTPQETL----- 317  
 Db 347 LLSFGELQYALSDKPEVVDPAVCCVTKYPTIYQPKYFLAESFASAKNKLKSWAATI 406  
 QY 318 ---FSIRH--FDELVELTSKL 333  
 Db 407 NRPFQIRNATQRVEILDKV 427

## RESULT 10

O96947  
 ID O96947 PRELIMINARY; PRT; 450 AA.  
 AC O96947;  
 DT 01-MAY-1999 (Tremblrel. 10, Created)  
 DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE PHENYLALANINE HYDROXYLASE.  
 OS Geodia cydonium (Sponge).  
 OC Eukaryota; Metazoa; Porifera; Demospongiae; Tetractinomorpha;  
 OC Astrophorida; Geodiidae; Geodia.  
 OX NCBI\_TaxID=6047;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE-99053257; PubMed-9877430;  
 RA Wiens M., Kozlowski C., Batel R., Mueller W.;  
 RT "Phenylalanine hydroxylase from the sponge Geodia cydonium:  
 implication for allorecognition and evolution of aromatic amino acid  
 hydroxylases.";

```
RL Dev. Comp. Immunol. 22:469-478 (1998).
DR EMBL: Y16353; CAA76184.1; -.
DR HSP: P00439; 4PAH.
DR InterPro: IPR001273; -.
DR InterPro: IPR001273; -.
DR Pfam: PF00351; bioterin_H; 1.
DR Pfam: PF01842; ACT; 1.
DR PRINTS: PR00372; FWYHDXRLASE.
DR SEQUENCE 450 AA; 51204 MW; 691880218BB9725D CRC64;

Query Match 9.4%; Score 178.5; DB 5; Length 450;
Best Local Similarity 25.8%; Pred. No. 6e-07;
Matches 55; Conservative 43; Mismatches 80; Indels 35; Gaps 6;

QY 107 WYRLSSRFLSKSCPRFLDYLEAFGLSDFLDH-----QAVIKFELETHFSY 158
D 186 WRTIFNLVDFPPTHACK---EHNHVPFLQENCGYREDNIPOLEEVSQYLSQCTGRLR 242
Q 159 PVSGFVAPHOYLSLLODRYFPFIASVMRTLDKDNFSLTDLHDLGHVWPLHPSSEFF 218
D 243 PVAGLLSRDFLAGLAFRVFHSQYIRHSQNYTPEDVCHLGHVCSV-ILFAQFS 301
QY 219 INMGRLFTKVEKQALPQSKQRIQIQLSNLAIVRCTFWFVESGLIENHGRKAYGAVL 278
D 302 QEIG-----LASGAPDEEYQQAL------LYWFTIEFLCKQDGQTRKAYGAGL 345
QY 279 ISSPOELGHAFIDNVRVLPLELQIIRLPFNTS 311
D 346 ISSFGLQYCLSDKPEVRPLD-----PFKTS 371

RESULT 11
Q9W0K2 PRELIMINARY; PRT; 555 AA.
AC Q9W0K2;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE CG9122 PROTEIN.
GN CG9122.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OC NCBI_TaxID=7227;
[1]
SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo J., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Fertiera S., Fleischmann W.,
RA Foeller C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.B., Harvey D., Helman J.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jaitani M., Koushan F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasako P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
```

Db 31 SRDQV-----WRNLSLGRHLADKAPVVLGEGLEATGIGSECIPSLDENNEKL 80  
 QY 151 LETHFSYVPVSGFVAPHOYLSLQDRFFPIASVMRTLDKDNFSLTPOLIHLGLHVPWLL 210  
 Db 81 ARLGWACVGRGFIPPAVFTELQAMGVLAADIRTHEIYETPAPDIVHESACHAPIIA 140  
 QY 211 HPFSFEFFINMGLFTKIEKV--QAL-----PSKKQRIQTQLQSNLIA-- 251  
 Db 141 NRYAEVLKACGLVGFKAISVEDQAVFEAIRNLVSVKEDPDASEEAAHAQARLEAASA 200  
 QY 252 -----IVRCFWFVESGLTENHEGKAYGAVLISSPOELGHAFIDNVRVLPLELD 301  
 Db 201 SRRVSESTRAGRLTWTAETGLGSGVASPRYAGLFTSIGEAQHCLTPAVKKPLSV- 259  
 QY 302 QIIRLPNFTQETLFSIRHDELVELTSKLEWML-----DOGLESI 345  
 Db 260 ACADMDYDITRMQQLFVARDQFHLFEVLAEFESTLSWREGDGLTEAL 309

RESULT 13  
 OI7498 PRELIMINARY; PRT; 438 AA.  
 OI7498;  
 01-JAN-1998 (TEMBLrel. 05, Created)  
 01-JAN-1998 (TEMBLrel. 05, Last sequence update)  
 01-MAR-2001 (TEMBLrel. 16, Last annotation update)  
 PHENYLALANINE HYDROXYLASE (EC 1.14.16.1).  
 PAH.  
 Branchiostoma floridae (Florida lancelet) (Amphioxus).  
 OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;  
 Branchiostoma.  
 OX NCBI\_TaxID=7739;  
 RN SEQUENCE FROM N.A.  
 RA Patton S.J., Luke G.N., Holland P.W.;  
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
 RMBL; AJ001677; CAA04917.1; -  
 DR HSSP; P04176; iph1; -  
 DR InterPro; IPR001273; -  
 DR InterPro; IPR002912; -  
 DR Pfam; PF00351; Dioplerin\_H; 2.  
 DR Pfam; PF01842; ACT; 1.  
 DR PROSITE; PS00046; HISTONE\_H2A; UNKNOWN\_1.  
 KW Oxidoreductase.  
 SQ SEQUENCE 438 AA; 49958 MW; 1619297DBDBF5EE7 CRC64;

Query Match 8.3%; Score 157; DB 5; Length 439;  
 Best Local Similarity 24.1%; Pred. No. 3.7e-05;  
 Matches 64; Conservative 37; Mismatches 129; Indels 36; Gaps 8;  
 QY 105 NLMYRLSSRSFLSKSKYCP-----FFLDYLEAFGLLSDFLDH-QAVIKFFELTHFSYYP 159  
 Db 170 NTRQFTLRKLYPTDACRHNYPVFLLMNCGFREDNIQLEDVSNFLKDCGTGFTLRP 229  
 QY 160 VSGFVAPHOYLSLQDRFFPIASVMRTLDKDNFSLTPOLIHLGLHVPWLLHPSFSEFFI 219  
 Db 230 VAGLLSSRDFLAGLAFVPHFTQYIRHHSKPLYTPDPVCHELLGHAPLADPSFAQFSQ 289  
 QY 220 NGRLETKVIEKQALPSKKRIQTQLQSNLIAIVRCFWFVESGLTENHEGKAYGAVLI 279  
 Db 290 EIG-----LASIGAPDFVVKLATL-----YWFTEFGLCGDGEVKAQRDCC 333  
 QY 280 SSPOELGHAFIDNVRVLPLELDQIIRLPNFTQETLFSIRHDELVELTSKLEWMLDQ 339  
 Db 334 HRSESCRYCLTDKPDIRFEPEKISVTKYITEYQPIYFVADSFODAKEVR--TWS--- 388  
 QY 340 GLLSEIPL-----YNOEKYLSGFEVL 360  
 Db 389 ---HSIPRPFVSHYN--PYTQSVIEL 409

RESULT 14  
 Q9SC23 PRELIMINARY; PRT; 1253 AA.  
 AC Q9SC23;  
 DT 01-MAY-2000 (TEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TEMBLrel. 16, Last annotation update)  
 DE DISEASE RESISTANCE-LIKE PROTEIN.  
 GN F26013.200.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;  
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;  
 OC Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN SEQUENCE FROM N.A.  
 RA Delsen M., Berger C., Cooke R., Grelllet F., Laudie M., Mewes H.W.,  
 RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RMBL; AL133452; CAB63020.1; -  
 DR InterPro; IPR000157; -  
 DR InterPro; IPR000767; -  
 DR InterPro; IPR001611; -  
 DR InterPro; IPR002182; -  
 DR Pfam; PF00560; LRR; 5.  
 DR Pfam; PF00931; NB-ARC; 1.  
 DR Pfam; PF01582; TIR; 1.  
 DR PRINTS; PR00364; DISEASERSIST.  
 DR SMART; SM00255; TIR; 1.  
 SQ SEQUENCE 1253 AA; 4211ADE0566C2B01 CRC64;

Query Match 5.8%; Score 109.5; DB 10; Length 1253;  
 Best Local Similarity 20.3%; Pred. No. 1.4; 130; Indels 127; Gaps 18;  
 Matches 81; Conservative 61; Mismatches 127; Indels 127; Gaps 18;  
 QY 77 VHLISLSKNQREGCS-----TDMVYVTPFFNNLWYLLSS-----RF 115  
 Db 421 IHL-----EGCGFFPRVEINVLVEKCLVSM-EGVVMHNLQSGRKLNGKRRS 472  
 QY 116 SLKSKYCPREFFDYLEAFG---LLSDFLDHQAV-----IKFEL-----ET 153  
 Db 473 RLKPLIKYFLEDRQVLGSEDIEAIFLDFPSALSFDVNPMAFENMYNRYLKICSSNPGN 532  
 QY 154 HFSYVPVSGFVAPHOYLSLQDRFFPIASVMRTLDKDNFSLTPOLIHLGLHVPWLLHPSFSEFFI 197  
 Db 533 HYALHLPKGVKSLPEELRLHWEHPPLSLPQDFNRLVILNMCYSKLORLWEGTKELG 592  
 QY 198 -LIHLLGHVPWLLHPSFSEFFIN-----GRL-----FTKVIE-----KVQ 233  
 Db 593 MLKRLMCHSQQLVGIEQLALNMEVIDLQGCARLQFLATGHFQHLRVINLSGCIKIK 652  
 QY 234 ALPSKKRIQTL---OSNLIAIVRCFWFVESGLTENHEGKAYGAVLISSPOELG-HAF 289  
 Db 653 SPFEVPPNIEELVYKGTGIRSIPTVTFSPQDNSFYDVKHDKFLNREVSDSQSLSIMVY 712  
 QY 290 IDNVRVLP-----LELQIIRLPNFT-----STQETLFSIRHDELVEL---TSKLEWM 336  
 Db 713 LDNLKVLDSQCLELDTQIGIPNKLKYLGTAKELPSLMHLSLVLDLNCNKLH 772  
 QY 337 LDQGL-----LESTPLYNQEKYLSG 356  
 Db 773 LPMGTGNLSSVLNLSGCGSELEDIQGIPRNLEELIYLAG 811

RESULT 15  
 Q9GF63 PRELIMINARY; PRT; 506 AA.  
 ID Q9GF63  
 AC Q9GF63;

DT	01-MAR-2001 (TreeBrel. 16, Created)	
DT	01-MAR-2001 (TreeBrel. 16, Last sequence update)	
DT	01-MAR-2001 (TreeBrel. 16, Last annotation update)	
DN	MATK	
GN	Arabis alpina (gray rockcress).	
OS	Chloroplast	
OC	Eukaryota, Viridiplantae, Embryophyta, Tracheophyta; Spermatophyta;	
OC	Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;	
OC	Brassicales; Brassicaceae; Arabis.	
ON	NCBI_TaxID=50452;	
OX	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=AALP2.	
RA	Koch M., Mitchell-Olds T.;	
RA	"Evolutionary analysis of plastidic maturase K and nuclear chalcone	
RP	synthase and their utility for phylogenetic reconstructions within the	
RP	Brassicaceae."	
RP	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.	
LN	EMBL; AF144329; AAC43298.1;	
KW	Chloroplast.	
SQ	SEQUENCE 506 AA; 60472 MW; 3BD5743D758BF23 CRC64;	

  

Query Match	5.6%;	Score 106.5;	DB 8;	Length 506;
Best Local Similarity	22.6%;	Pred. No. 0.75;		
Matches 68;	Conservative 37;	Mismatches 131;	Indels 65;	Gaps

  

OY	3	YCE-RTLDPRYILKALKUQSLFFQNSQSLQRAYSTPYSVYRILOENKEKQALR 61
Db	190	YCNWKNFD---IKKKLIINPRFFLYNSHVCE-----YESIFFLKRSSHLSRTAY 239
OY	62	HKCSISLEFFKNL-----LFVH-----LISLKNQ-----REGCSTDMAVSTP----- 100
Db	240	EVLPERILFYAKTOHFLKVFVNPNFALGLGLKDPFLHYRVYHGKSILATKDTPLLNKNK 299
OY	101	PFNNLWRLLSRFSFLWKSXCPFFLDYLEAFGLSLDFLDHQAVKFFFELETHFSYVPV 160
Db	300	FYFNLMOFFYSVWQSQKHNLGSKDLNLEFGLYLSLSRLNPLVWVSQMLENSFLI--- 356
OY	161	SGFVAPHOYLILQDRYPPIASVMTKLDKNFSLTPDLIHLGLH-----VPVLLHFSFSE 216
Db	357	-----DNIRIKLONKIPISIGISGTSKDKFC-----NLHGHPISKANW-TESSDSD 401
OY	217	FFINMGRLETKVIERQVALPSKKOR-----IOTLQSNLIATVRCFWFTVESGLI 265
Db	402	ILNRFVRCRNISHYSQSSKKHLRIKYLRLCCVTKLARKHKSTVAFRLKRLGSL 461
OY	266	E 266
Db	462	E 462

Search completed: June 29, 2001, 13:55:27  
Job time: 411 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 29, 2001, 13:40:51 ; Search time 40.35 Seconds  
(without alignments)  
683.400 Million cell updates/sec

Title: US-09-438-185-1047

Perfect score: 1889

Sequence: 1 VHCERTLDPKYLKIALK.....ESTIPLYNQKYLSGFEVLQ 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 segs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Database DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR.68.\*

1: PIR1.\*

2: PIR2.\*

3: PIR3.\*

4: PIR4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1886	99.8	362	2 E72002	aromatic amino aci
2	1886	99.8	362	2 C86621	aromatic amino aci
3	256.5	13.6	289	2 D82413	phenylalanine-4-hy
4	235.5	12.5	262	2 F83535	phenylalanine-4-hy
5	231.5	12.3	444	1 S10489	tryptophan 5-monoo
6	229.5	12.1	262	2 A53452	phenylalanine hydr
7	229.5	12.1	444	2 S51159	tryptophan 5-monoo
8	228.5	12.1	491	2 A28582	tyrosine 3-monooxy
9	227.5	12.0	444	1 WHRTW	tryptophan 5-monoo
10	226.5	12.0	447	2 A34582	tryptophan 5-monoo
11	225.5	11.9	498	2 JX0068	tyrosine 3-monooxy
12	223.5	11.8	481	2 I31367	tryptophan 5-monoo
13	223	11.8	498	1 WHRT1	tyrosine 3-monooxy
14	221.5	11.7	491	2 I45983	tyrosine 3-monooxy
15	220	11.6	528	1 WHH04	tyrosine 3-monooxy
16	219	11.6	579	1 A55369	tyrosine 3-monooxy
17	214	11.3	453	1 WHRTF	phenylalanine 4-mo
18	213.5	11.3	452	1 WHRTF	phenylalanine 4-mo
19	212.5	11.2	453	2 J10039	phenylalanine 4-mo
20	211.5	11.2	491	2 J10039	phenylalanine 4-mo
21	207	11.0	453	2 J00766	phenylalanine 4-mo
22	206.5	10.9	575	2 T34509	hypothetical prote
23	197	10.4	453	1 A42271	tryptophan 5-monoo
24	194	10.3	452	2 J04888	phenylalanine 4-mo
25	178.5	9.4	404	2 T25453	tyrosine 3-monooxy
26	175.5	9.3	457	2 T23494	phenylalanine 4-mo
27	109.5	5.8	1253	2 T45787	disease resistance
28	108.5	5.7	1501	1 B29813	174K ninaC protein
29	106.5	5.6	296	2 A40996	phenylalanine 4-mo

30	104.5	5.5	2331	2 S44054	genome polyprotein
31	102	5.4	563	2 T38766	probable transcrip
32	102	5.4	1050	2 C81624	exodeoxyribonuclea
33	100.5	5.3	1157	2 S38160	NUP133 protein - y
34	99	5.2	924	2 T09220	exocyst complex pr
35	98.5	5.2	584	2 D81265	hypothetical prote
36	98.5	5.2	1024	2 C84208	hypothetical prote
37	98	5.2	1846	2 T33079	hypothetical prote
38	97	5.1	492	1 S68856	cytochrome P450 2L
39	97	5.1	1050	2 G86582	exodeoxyribonuclea
40	96.5	5.1	512	2 G86520	glucose-6-P dehydro
41	96.5	5.1	512	2 C72103	glucose-6-P dehydro
42	96	5.1	529	2 S62468	probable membrane
43	96	5.1	1050	2 H72041	exodeoxyribonuclea
44	96	5.1	2059	2 T41933	large tegument pro
45	95.5	5.1	536	2 E70066	hypothetical prote

ALIGNMENTS

RESULT 1

E72002 aromatic amino acid hydroxylase - Chlamydomophila pneumoniae (strain CWL029)

C:Species: Chlamydomophila pneumoniae; Chlamydia pneumoniae

C:Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 05-May-2000

C:Accession: E72002

R:Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood,

Nature Genet. 21, 385-389, 1999

A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A:Reference number: A72000; MUID:99206606

A:Accession: E72002

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-362 <ARN>

A:Cross-references: GB:AE001685; GB:AE001363; NID:94377378; PIDN:AA19183.1; PID:9437

A:Experimental source: strain CWL029

C:Genetics:

A:Gene: CPn1046

Query Match 99.8% Score 1886; DB 2; Length 362;  
Best Local Similarity 99.7%; Pred. No. 1.3e-144;  
Matches 361; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VHCERTLDPKYLKIALKQSLFFQNSQSLQRAYSTPYSYRILLOKENEKQALA 60

Db 1 MHYCERTLDPKYLKIALKQSLFFQNSQSLQRAYSTPYSYRILLOKENEKQALA 60

Qy 61 RHKCSILEFFKNLLFVHLLSLSKNQREGGCTDMVVSTPFNNRLWYLLSSRFLSKWS 120

Db 61 RHKCSILEFFKNLLFVHLLSLSKNQREGGCTDMVVSTPFNNRLWYLLSSRFLSKWS 120

Qy 121 YCPRFLLDYLEAFGLLSDFLDHQAVIKFFFELETHFSYYPVSGFVAPHOYLSLQDRFPPI 180

Db 121 YCPRFLLDYLEAFGLLSDFLDHQAVIKFFFELETHFSYYPVSGFVAPHOYLSLQDRFPPI 180

Qy 181 ASVMRTLDKDNFSLTPDLHLLGHVFWLLHPSFSEFFINMGRLFTKVIKQVQALPSKKQ 240

Db 181 ASVMRTLDKDNFSLTPDLHLLGHVFWLLHPSFSEFFINMGRLFTKVIKQVQALPSKKQ 240

Qy 241 RIQTQSLNLAIVRCFFWTFESGLTENHEGRKAYGAVLISSPOELGHAFIDNVRVLPLEL 300

Db 241 RIQTQSLNLAIVRCFFWTFESGLTENHEGRKAYGAVLISSPOELGHAFIDNVRVLPLEL 300

Qy 301 DOIIRPFNTSTPQETLSIRHDFELVELTSKLEWMLDQGLLESIPLYNOEKYLSGFEVL 360

Db 301 DOIIRPFNTSTPQETLSIRHDFELVELTSKLEWMLDQGLLESIPLYNOEKYLSGFEVL 360

Qy 361 CQ 362

Db 361 CQ 362

## RESULT 2

C86621  
aromatic amino acid hydroxylase [imported] - Chlamydothrix pneumoniae (strain J138)  
C:Species: Chlamydothrix pneumoniae, Chlamydothrix pneumoniae  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 02-Mar-2001  
C:Accession: C86621  
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise  
Nucleic Acids Res. 28, 2311-2314, 2000  
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.  
A:Reference number: A86491; MUID:20330349  
A:Accession: C86621  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-362 <STO>  
A:Cross-references: GB:BA000008; NID:g8979419; PIDN:BAA99253.1; GSPDB:GN00142  
A:Experimental source: strain J138  
C:Genetics:  
A:Gene: CP1046

Query Match 99.8%; Score 1886; DB 2; Length 362;  
Best Local Similarity 99.7%; Pred. No. 1.3e-144; Indels 0; Gaps 0;  
Matches 361; Conservative 1; Mismatches 0;

QY 1 VHYCERTLPKYLKALKURQSLFFQNSQSQRAYSTPYSYRILOKQENKQKALA 60  
DB 1 MHYCERTLPKYLKALKURQSLFFQNSQSQRAYSTPYSYRILOKQENKQKALA 60  
QY 61 RHKCTISILEFFKNLLFVHLLSKNORECGCTDMAVSTPFFNRLMYRLSSRFLWKS 120  
DB 61 RHKCTISILEFFKNLLFVHLLSKNORECGCTDMAVSTPFFNRLMYRLSSRFLWKS 120  
QY 121 YCPREFDYLAEAGLLSDFLDHOAVIKFPELTHFSYYPVSGFVAPHOYLSLLQDRYFPI 180  
DB 121 YCPREFDYLAEAGLLSDFLDHOAVIKFPELTHFSYYPVSGFVAPHOYLSLLQDRYFPI 180  
QY 181 ASVMTLDKDNFSLTPDLIHDLGHVPMLLHPSSEFFINMGRFTKVIKVOALPSKKO 240  
DB 181 ASVMTLDKDNFSLTPDLIHDLGHVPMLLHPSSEFFINMGRFTKVIKVOALPSKKO 240  
QY 241 RIQTQSLNLAIVRCFWTFVESGLIENHGRKAYGAVLISSPQELGHAFIDNVRVLPLEL 300  
DB 241 RIQTQSLNLAIVRCFWTFVESGLIENHGRKAYGAVLISSPQELGHAFIDNVRVLPLEL 300  
QY 301 DQIIRLPNTSTPQETLSIRHFDDELVELTSKLEWMLDQGLLESIPLYNOEKYLSGFEVL 360  
DB 301 DQIIRLPNTSTPQETLSIRHFDDELVELTSKLEWMLDQGLLESIPLYNOEKYLSGFEVL 360  
QY 361 CQ 362  
DB 361 CQ 362

RESULT 3  
Phenylalanine-4-hydroxylase VCA0828 [imported] - Vibrio cholerae (strain N16961 serogroup  
C:Species: Vibrio cholerae  
C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: D82413  
R:Heldberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
Hardison, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B.  
I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: D82413  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-289 <HEI>  
A:Cross-references: GB:AE004410; GB:AE003853; NID:g9658244; PIDN:AAF96726.1; GSPDB:GN001  
A:Experimental source: serogroup O1; strain N16961; biotype El Tor  
C:Genetics:

A:Gene: VCA0828  
A:Map position: 2

Query Match 13.6%; Score 256.5; DB 2; Length 289;  
Best Local Similarity 25.6%; Pred. No. 2.4e-13;  
Matches 65; Conservative 56; Mismatches 108; Indels 25; Gaps 6;  
QY 106 LWYRLSSRFLSKSYCPREFDYLAEAGLLSDFLDHOAVI-KFEELETHFSYYPVSGFV 164  
DB 49 VVHELITRQEQVAVTRACQAYLDGLANLNLPTDLRQLPQLEINRVLQRETGHQVPEVPAI 108  
QY 165 APHOYLSLLQDRYPIASVMRTLDKDNFSLTPDLIHDLGHVPMLLHPSSEFFINMGR 224  
DB 109 SPDRFALLADKKEPVATFELRREEDYLPQEPDFHVEYVGHGCMALTHPDEAFATHTVYQ 168  
QY 225 FTKVIEKVOALPSKKQRIQTQLSGLIENHGRKAYGAVLISSPQEL 284  
DB 169 GAKATPKERSY-----LRLYFVTFVGLVQEQGTIYGGGILSSPGE 212  
QY 285 LGHAFIDNV-RVLPLELDQIIRLPNTSTPQETLSIRHFDDELVELTSK----LEWM-LD 338  
DB 213 TLYASESTIPREFDLMQVLRTPYRIDIMQPIYVLPDLSOLYSORDYVNLVWQAMQ 272  
QY 339 QGLLESIPLYNOEK 352  
DB 273 DGLLP--PLFQPK 284

RESULT 4  
F83535  
Phenylalanine-4-hydroxylase PA0872 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: F83535  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;  
Adnan, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L.  
; Lory, S.; Olson, M.V.  
Nature 406, 959-964, 2000  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa  
A:Reference number: A82950; MUID:20437337  
A:Accession: F83535  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-262 <STO>  
A:Cross-references: GB:AE004522; GB:AE004091; NID:g9946768; PIDN:AAG04261.1; GSPDB:GN  
A:Experimental source: strain PA01  
C:Genetics:  
A:Gene: phhA; PA0872

Query Match 12.5%; Score 235.5; DB 2; Length 262;  
Best Local Similarity 24.5%; Pred. No. 1e-11;  
Matches 61; Conservative 52; Mismatches 103; Indels 33; Gaps 6;  
QY 93 DMAVYSTPFFNRLMYRLSSRFLSKSYCPREFDYLAEAGLLSDFLDHOAVIKFPELE 152  
DB 13 DNGFHYETHEQVWNLITRLQKLVIEGRACQYLDGIEQLG-----LPHERIPQLDEIN 67  
QY 153 -----THFSYYPVSGFVAPHOYLSLLQDRYPIASVMRTLDKDNFSLTPDLIHDLGHV 206  
DB 68 RVLQATTCGRVARVPALIPFQTFPELLSQQFPVATFHTPELDYLPQEPDIFHEIFGHC 127  
QY 207 PWLLHPSSEFFINMGRFTKVIKVOALPSKKQRIQTQLSGLIENHGRKAYGAVLISSPQEL 266  
DB 128 PLTNPWFPAEFTHTYGLGLKA-----SKEER-----VFLARLYWMTTEFGIVE 171  
QY 267 NHEGRKAYGAVLISSPQELGHAFID--NVRVLPLELDQIIRLPNTSTPQETLSIRHF 323  
DB 172 TDGKRIYGGGILSSPKETVYLSLDEPLHQAFLNPLE---AMRTPYRIDILPLYFLVLPDL 228  
QY 324 DELVELTSK 332

Db 229 KRLFOLAQE 237

RESULT 5

S10489

tryptophan 5-monooxygenase (EC 1.14.16.4) - human

N:Alternate names: tryptophan 5-hydroxylase

C:Species: Homo sapiens (man)

C>Date: 04-Sep-1998 #sequence\_revision 04-Sep-1998 #text\_change 03-Mar-2000

C:Accession: S10489; S51559

R:Boularand, S.; Darmon, M.C.; Ganem, Y.; Launay, J.M.; Mallet, J.

Nucleic Acids Res. 18, 4257, 1990

A:Title: Complete coding sequence of human tryptophan hydroxylase.

A:Reference number: S10489; MUID:90332431

A:Accession: S10489

A:Molecule type: mRNA

A:Residues: 1-444 <BOU>

A:Cross-references: EMBL:X52836; NID:937954; PIDN:CAA37018.1; PID:937955

R:Tipper, J.P.; Citron, B.A.; Ribeiro, P.; Kaufman, S.

Arch. Biochem. Biophys. 315, 445-453, 1994

A:Title: Cloning and expression of rabbit and human brain tryptophan hydroxylase cDNA in

Reference number: S51199; MUID:95077422

Accession: S51559

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-18, 'T', '20-67, 'T', '69-89, 'TP', '92-96, 'M', '98-99, 'E', '101-103, 'S', '105-150, 'S', '15

35, 'G', '437-444 <TIP>

A:Cross-references: GB:L29306; NID:9531192; PIDN:AAA67050.1; PID:9531193

C:Genetics:

A:Gene: GDB:TPH; TPRH

A:Cross-references: GDB:120732; OMIM:191060

A:Map position: 11p15.1-11p14.3

C:Function:

A:Description: catalyzes the oxidation of tryptophan to 5'-hydroxytryptophan by tetrahyd

A:Pathway: melatonin biosynthesis; serotonin biosynthesis; tryptophan catabolism

C:Superfamily: phenylalanine 4-monooxygenase

C:Keywords: biotin; iron; melatonin biosynthesis; metalloprotein; monooxygenase; oxid

F:58/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status predict

F:260,443/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #sta

F:272,277,317/Binding site: iron (His, His, Glu) #status predicted

Query Match 12.3%; Score 231.5; DB 1; Length 444;

Best Local Similarity 27.8%; Pred. No. 4.4e-11;

Matches 64; Conservative 43; Mismatches 96; Indels 27; Gaps 4;

QY 107 WYRLSSRFSILWKSYPFRFFLDYLEAFGLLSDFLDH-----QAVIKFEELETHFSY 158

Db 174 WGTVPQELNKLYPHACR---EYLNKPLLSKYCYREDNIPQEDISNFKLRTGFSIR 230

159 PVSGVAPHOVLSLQDRYPFIASVMRTLDKDNFSLTPDLIHDLGHVPLWLLHPSFSEFF 218

Db 231 PVAGYLSPRDELGLAFRVFCHQVYVHSSDPYTPEDTCHELLGHVPLLAAPSFAQFS 290

QY 219 INMGRLFTKVIEKVALPSKKRIOTLQSNLIAIVRCFWFTVSGLIENHEGKAYGAVL 278

Db 291 QEIG-----LASLGASEAVQKLA-----CYFFIVFGLCKDQGLRVFAGL 334

QY 279 ISGPQELGHADINRVLPLELQIIRLPNTSTPOETLFSIRHDELVE 328

Db 335 LSSISLKHLSGHAHVFPDPKPKCKQECILITTFQDYVFSSESFEAKE 384

RESULT 6

A53452

phenylalanine hydroxylase - Pseudomonas aeruginosa

C:Species: Pseudomonas aeruginosa

C>Date: 26-Jul-1996 #sequence\_revision 26-Jul-1996 #text\_change 08-Oct-1999

C:Accession: A53452

R:Zhao, G.S.; Xia, T.; Song, J.; Roy, R.A.

Proc. Natl. Acad. Sci. U.S.A. 91, 1366-1370, 1994

A:Title: Pseudomonas aeruginosa possesses homologues of mammalian phenylalanine hydroxyl

A:Reference number: A53452; MUID:94151331

A:Accession: A53452

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-262 <RES>

A:Cross-references: GB:M88627; NID:9476740; PIDN:AAA25936.1; PID:9476741

Query Match 12.1%; Score 229.5; DB 2; Length 262;

Best Local Similarity 24.1%; Pred. No. 3.2e-11;

Matches 60; Conservative 52; Mismatches 104; Indels 33; Gaps 6;

QY 93 DMAVYSTFFNRLWYRLSSRFSILWKSYPFRFFLDYLEAFGLLSDFLDHQAIVKFELE 152

Db 13 DNGFIRYETEHQVWNTLITROLKVIEGRACQEQYLDGIEQLG-----LPHRIPOLDEN 67

QY 153 -----THFSYYPVSGVAPHOVLSLQDRYPFIASVMRTLDKDNFSLTPDLIHDLGHV 206

Db 68 RVLOATTGWRVARVPALIPQTFEFLASQCFPVATFRTPEELDYLOQEPDIFHEIFGHC 127

QY 207 PWLLHPSFSEFFINMGRLFTKVIEKVALPSKKRIOTLQSNLIAIVRCFWFTVSGSLIE 266

Db 128 PLITNPWLAEFTHTYGLGLKA-----SKEER-----VFLARLYWMTIEFGLVE 171

QY 267 NHEGRKAYCAVLISSPQELGHAFID---NVRVLPLELQIIRLPNTSTPOETLFSIRHF 323

Db 172 TDGKRRIYGGGILSPKRYVLSDEPLHQAENPLE---AMRTPYRIDILQPLFYFLPDL 228

QY 324 DELVELTSK 332

Db 229 KRLFOLAQE 237

RESULT 7

S51199

tryptophan 5-monooxygenase (EC 1.14.16.4) - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 15-Jul-1995 #sequence\_revision 19-Apr-1996 #text\_change 31-Mar-2000

C:Accession: S51199; A32699

R:Tipper, J.P.; Citron, B.A.; Ribeiro, P.; Kaufman, S.

Arch. Biochem. Biophys. 315, 445-453, 1994

A:Title: Cloning and expression of rabbit and human brain tryptophan hydroxylase cDNA

A:Reference number: S51199; MUID:95077422

A:Accession: S51199

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-444 <TIP>

A:Cross-references: NID:9531213; PIDN:AAA67051.1; PID:9531214

R:Greenett, H.E.; Ledley, F.D.; Reed, L.L.; Woo, S.L.C.

Proc. Natl. Acad. Sci. U.S.A. 84, 5530-5534, 1987

A:Title: Full-length cDNA for rabbit tryptophan hydroxylase: functional domains and e

A:Reference number: A32699; MUID:87289638

A:Accession: A32699

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-101, 'L', '103-150, 'L', '152-201, 'ND', '204-206, 'R', '208-389, 'K', '391-444 <GRE>

A:Cross-references: GB:M17250; NID:9165771; PIDN:AAA31487.1; PID:9165772

C:Superfamily: phenylalanine 4-monooxygenase

C:Keywords: biotin; iron; metalloprotein; oxidoreductase; phosphoprotein

F:272,317/Binding site: iron (His, His, Glu) #status predicted

Query Match 12.1%; Score 229.5; DB 2; Length 444;

Best Local Similarity 27.4%; Pred. No. 6.4e-11;

Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;

QY 107 WYRLSSRFSILWKSYPFRFFLDYLEAFGLLSDFLDHQA-----VIKFELETHFSY 158

Db 174 WGTVPQELNKLYPHACR---EYLNKPLLSKYCYREDNIPQEDISNFKLRTGFSIR 230

QY 159 PVSGVAPHOVLSLQDRYPFIASVMRTLDKDNFSLTPDLIHDLGHVPLWLLHPSFSEFF 218

Db 231 PVAGYLSPRDELGLAFRVFCHQVYVHSSDPYTPEDTCHELLGHVPLLAAPSFAQFS 290

```
QY 219 INNGRLFTKVIKVOALPSKKQRIOTLQSNLIAIVRCFWFTVESGLIENHEGRKAYGAVL 278
Db 291 QEIG-----LASLGAEAAVQLAT-----CYFTTVEFLGCKQDQQLRVFGAGL 334
QY 279 ISSPOELGHAFIDNVRVLPLELDQIIRLPFNSTPQETLFSIRHFDELVE 328
Db 335 LSSISLHSLHSGHAKVPDPKITYKQECLETTTQDVFVFSSEFDEAKE 384

RESULT 8
A28582
A:Title: tyrosine 3-monooxygenase (EC 1.14.16.2) - quail
N:Alternate names: tyrosine 3-hydroxylase
C:Species: Phasianidae gen. sp. (quail)
C:Date: 19-Nov-1988 #sequence_revision 19-Nov-1988 #text_change 31-Mar-2000
C:Accession: A28582; PH1524
R:Pauquet, M.; Grima, B.; Lamouroux, A.; Mallet, J.
J. Neurochem. 50, 142-148, 1988
A:Title: Cloning of quail tyrosine hydroxylase: amino acid homology with other hydroxylase
A:Reference number: A28582; MUID:88089590
A:Accession: A28582
A:Molecule type: mRNA
A:Residues: 1-491 <FAU>
A:Cross-references: GB:M24778; NID:G213649; PIDN:AAA49514.1; PID:G213650
R:Pauquet, M.; Boni, C.
J. Neurochem. 60, 274-281, 1993
A:Title: The quail tyrosine hydroxylase gene promoter contains an active cyclic AMP-resp
A:Reference number: PH1524; MUID:93107923
A:Accession: PH1524
A:Molecule type: DNA
A:Residues: 1-30 <FA2>
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bioprotein; catecholamine biosynthesis; iron; metalloprotein; oxidoreductase;
F:324,329,369/Binding site: iron (His, His, Glu) #status predicted

Query Match 12.1%; Score 229.5; DB 2; Length 491;
Best Local Similarity 29.1%; Pred. No. 7.3e-11;
Matches 66; Conservative 34; Mismatches 100; Indels 27; Gaps 4;

QY 107 WYRLSSRFLSKYSCPRFFLDYLEAFGLLSDFLDH-----QAVIKFELETHFSY 158
Db 226 MREVTSLKSLYPTHACK---EYLEAFNLEKFCGYNNENIPQLEEYSRFLKERTGQFLR 282
QY 159 PYSGVAPHOYLSLQDRYPPIASVMTLKDKNFSLTDLHLLGHVPLLLHPSSEFF 218
1 283 PYRGLLSARDELASLAFRVQCQYTRHASSPMHSPEDPCCHLLGHVPLMLADKTEAQS 342
QY 219 INNGRLFTKVIKVOALPSKKQRIOTLQSNLIAIVRCFWFTVESGLIENHEGRKAYGAVL 278
Db 343 QDIG-----LASLGATDEIEKIALT-----YFTTVEFLGCRQNGIVKAYGAGL 386
QY 279 ISSPOELGHAFIDNVRVLPLELDQIIRLPFNSTPQETLFSIRHFDE 325
Db 387 LSSYELHLSLSEPEVRDFPDAAAVQCQDPQYQPVYFVSEFSFSD 433

RESULT 9
WHRTW
A:Title: tyrosine 3-monooxygenase (EC 1.14.16.4) - rat
N:Alternate names: tyrosine 3-hydroxylase
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 03-Mar-2000
C:Accession: J10034; A60034; A24367
R:Darmon, M.C.; Guilbert, B.; Leviel, V.; Ehret, M.; Mallet, J.
J. Neurochem. 51, 312-316, 1988
A:Title: Sequence of two mRNAs encoding active rat tyrosine hydroxylase.
A:Reference number: J10034; MUID:88244702
A:Accession: J10034
A:Molecule type: mRNA
A:Residues: 1-444 <DAR>
A:Cross-references: GB:X53501; NID:957760; PIDN:CAA37579.1; PID:g57761
A:Experimental source: pineal gland
```

```
R:Kim, K.S.; Wessel, T.C.; Stone, D.M.; Carver, C.H.; Joh, T.H.; Park, D.H.
Brain Res. Mol. Brain Res. 9, 277-283, 1991
A:Title: Molecular cloning and characterization of cDNA encoding tryptophan hydroxyla
A:Reference number: A60034; MUID:91245924
A:Accession: A60034
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-444 <KIM>
A:Experimental source: dorsal raphe nucleus
R:Darmon, M.C.; Grima, B.; Cash, C.D.; Maitre, M.; Mallet, J.
FEBS Lett. 206, 43-46, 1986
A:Title: Isolation of a rat pineal gland cDNA clone homologous to tyrosine and phenyl
A:Reference number: A24367; MUID:87005247
A:Accession: A24367
A:Molecule type: mRNA
A:Residues: 167-261 <DAR>
A:Cross-references: GB:M28000; NID:G207432; PIDN:AAA42262.1; PID:g207433
C:Comment: This enzyme has different physical properties in pineal gland and in dorsa
C:Function:
A:Description: catalyzes the oxidation of tryptophan to 5'-hydroxytryptophan by tetra
A:Pathway: melatonin biosynthesis; serotonin biosynthesis; tryptophan catabolism
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bioprotein; iron; melatonin biosynthesis; metalloprotein; monooxygenase; o
F:58/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status pred
F:260,443/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #
F:272,277,317/Binding site: iron (His, His, Glu) #status predicted

Query Match 12.0%; Score 227.5; DB 1; Length 444;
Best Local Similarity 27.4%; Pred. No. 9.3e-11;
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;

QY 107 WYRLSSRFLSKYSCPRFFLDYLEAFGLLSDFLDH-----QAVIKFELETHFSY 158
Db 174 WGTIFRELKLYPTHACK---EYLRNLPILLSKYCYREDNVNPOLEDVSNFLKERTGSIR 230
QY 159 PYSGVAPHOYLSLQDRYPPIASVMTLKDKNFSLTDLHLLGHVPLLLHPSSEFF 218
Db 231 PVAGLSLPRDFSLGLAFRVFCHTQYVRHSSDPLTYPTPEPTCHELLGHVPLLAEPSPAQS 290
QY 219 INNGRLFTKVIKVOALPSKKQRIOTLQSNLIAIVRCFWFTVESGLIENHEGRKAYGAVL 278
Db 291 QBIG-----LASLGAEETVOKLAT-----CYFTTVEFLGCKQDQQLRVFGAGL 334
QY 279 ISSPOELGHAFIDNVRVLPLELDQIIRLPFNSTPQETLFSIRHFDELVE 328
Db 335 LSSISLHSLHSGHAKVPDPKVPACKQECLETSFDQVYFVSESFDEAKE 384

RESULT 10
A34582
A:Title: tyrosine 3-monooxygenase (EC 1.14.16.4) - mouse
C:Species: Mus musculus (house mouse)
C:Date: 22-Jun-1990 #sequence_revision 09-Oct-1992 #text_change 31-Mar-2000
C:Accession: A34582
R:Stoll, J.; Kozak, C.A.; Goldman, D.
Genomics 7, 88-96, 1990
A:Title: Characterization and chromosomal mapping of a cDNA encoding tryptophan hydro
A:Reference number: A34582; MUID:90243261
A:Accession: A34582
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-447 <STO>
A:Cross-references: GB:J04758; NID:G202113; PIDN:AAA63401.1; PID:g202114
A:Note: the authors translated the codon AAC for residue 405 as Gln
C:Superfamily: phenylalanine 4-monooxygenase
C:Keywords: bioprotein; iron; metalloprotein; oxidoreductase; phosphoprotein
F:275,280,320/Binding site: iron (His, His, Glu) #status predicted

Query Match 12.0%; Score 226.5; DB 2; Length 447;
Best Local Similarity 27.4%; Pred. No. 1.1e-10;
Matches 63; Conservative 43; Mismatches 97; Indels 27; Gaps 4;
```



Qy 107 WYLLSSRFLSKYCPREFLDYLAAGLLSDFLDH-----QAVIKFELETHFSY 158  
 Db 177 WGTIVRELKLYPHACR---EYLNPLLSKYCYREDNIPQLEDVSNELKERTGFSIR 233  
 Qy 159 PVSGFVAPHOYLSLQDRIYPIASVMTLDKFNLSLTPDLHDLGHVPMWLLHPSSEFF 218  
 Db 234 PVAGYLSRDFLGLAFVFCITQYVHSSDPLTPEDPTCHELLGHVPLLAEPSPAQS 293  
 Qy 219 INMGRLFTKVIKVOALPSKKORIOTLOSMLIAIVRCFWFTVESGLTENHEGKAYCAVL 278  
 Db 294 QEIG-----LASLGASEETVKLAT-----CYFTVEFGLCQDCQLRVFAGL 337  
 Qy 279 ISSPQELGHAFIDNVRVLPFLDQIIRLPNTSTPQETLFSIRHDELVE 328  
 Db 338 LSSISELKHALSHAKVRPFDPFKJACKOECLITSFQDVYFVSESEDAKE 387

RESULT 11  
 JN0068  
 tyrosine 3-monooxygenase (EC 1.14.16.2) - mouse  
 A:Alternate names: tyrosine hydroxylase  
 C:Species: Mus musculus (house mouse)  
 C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 31-Mar-2000  
 C:Accession: JN0068; S21322  
 R:Chikawa, S.; Sasaoka, T.; Nagatsu, T.  
 Biochem. Biophys. Res. Commun. 176, 1610-1616, 1991  
 A:Title: Primary structure of mouse tyrosine hydroxylase deduced from its cDNA.  
 A:Reference number: JN0068; MUID:91248263  
 A:Accession: JN0068  
 A:Molecule type: mRNA  
 A:Residues: 1-498 <ICH>  
 A:Cross-references: GB:M69200; MID:g201997; PIDN:AAA40434.1; PID:g201998  
 A:Experimental source: brain  
 R:Morgan, W.W.; Bermudez, J.; Sharp, Z.D.  
 submitted to the EMBL Data Library, June 1990  
 A:Description: DC-12 Nuclear Extracts Produce Tissue-Specific Protection of Several Sequences  
 A:Reference number: S21322  
 A:Accession: S21322  
 A:Molecule type: DNA  
 A:Residues: 1-30 <MOR>  
 A:Cross-references: EMBL:X53503; MID:g55055; PIDN:CAA37580.1; PID:g55056  
 C:Comment: This enzyme, which requires ferrous iron, catalyzes the hydroxylation of tyrosine in the physiology of adrenergic neurones.  
 C:Superfamily: phenylalanine 4-monooxygenase  
 C:Keywords: biotin; iron; metalloprotein; monooxygenase; oxidoreductase; phosphoprotein; 331,336,376/Binding site: iron (His, His, Glu) #status predicted  
 F:331,336,376/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.8%; Score 225.5; DB 2; Length 498;  
 Best Local Similarity 27.5%; Pred. No. 1.6e-10;  
 Matches 66; Conservative 42; Mismatches 101; Indels 31; Gaps 5;

Qy 107 WYLLSSRFLSKYCPREFLDYLAAGLLSDFLDH-----QAVIKFELETHFSY 158  
 Db 233 WKEVYATLKLTAHACR---EHLERFQLLYCYREDNIPQLEDVSHFLKERTGQLR 289  
 Qy 159 PVSGFVAPHOYLSLQDRIYPIASVMTLDKFNLSLTPDLHDLGHVPMWLLHPSSEFF 218  
 Db 290 PVAGLSARDFLGLAFVFCITQYVHSSDPLTPEDPTCHELLGHVPLLAEPSPAQS 349  
 Qy 219 INMGRLFTKVIKVOALPSKKORIOTLOSMLIAIVRCFWFTVESGLTENHEGKAYCAVL 278  
 Db 350 QDIG-----LASLGASEETVKLAT-----CYFTVEFGLCQDCQLRVFAGL 393  
 Qy 279 ISSPQELGHAFIDNVRVLPFLDQIIRLPNTSTPQETLFSIRHDELVE 328  
 Db 394 LSSYGLLSLSEEPVAFDPDAAVQPYQDQYFVSESDAKDLRNYSRIQ 453

RESULT 12  
 JN1567  
 tryptophan 5-monooxygenase (EC 1.14.16.4) - African clawed frog

C:Species: Xenopus laevis (African clawed frog)  
 C:Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 31-Mar-2000  
 C:Accession: JN1567  
 R:Green, C.B.; Besharse, J.C.  
 J. Neurochem. 62, 2420-2428, 1994  
 A:Title: Tryptophan hydroxylase expression is regulated by a circadian clock in Xenopus  
 A:Reference number: JN1567; MUID:94246419  
 A:Accession: JN1567  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-481 <GRE>  
 A:Cross-references: GB:I20679; MID:g450644; PIDN:AAA21306.1; PID:g450645  
 C:Superfamily: phenylalanine 4-monooxygenase  
 C:Keywords: biotin; iron; metalloprotein; oxidoreductase  
 F:309,314,354/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.8%; Score 223.5; DB 2; Length 481;  
 Best Local Similarity 29.5%; Pred. No. 2.2e-10;  
 Matches 70; Conservative 38; Mismatches 90; Indels 39; Gaps 9;

Qy 107 WYLLSSRFLSKYCPREFLDYLAAGLLSDFLDH-----QAVIKFELETHFSY 158  
 Db 211 WGTIVRELKLYPHACR---EYLNPLLSKYCYREDNIPQLEDVSNELKERTGFSIR 267  
 Qy 159 PVSGFVAPHOYLSLQDRIYPIASVMTLDKFNLSLTPDLHDLGHVPMWLLHPSSEFF 217  
 Db 268 PVAGYLSRDFLGLAFVFCITQYVHSSDPLTPEDPTCHELLGHVPLLAEPSPAQS 326  
 Qy 218 INMGRLFTKVIKVOALPSKKORIOTLOSMLIAIVRCFWFTVESGLTENHEGKAYCAVL 277  
 Db 327 SQEIG-----LASLGASEETVKLAT-----CYFTVEFGLCQDCQLRVFAGL 370  
 Qy 278 LSSPQELGHAFIDNVRVLPFLDQIIRLPNTSTPQETLFSIRHDELVE 334  
 Db 371 LSSISELKHALSHAKVRPFDPFKJACKOECLITSFQDVYFVSESEDAKE 417

RESULT 13  
 WHRYT  
 tyrosine 3-monooxygenase (EC 1.14.16.2) - rat  
 A:Alternate names: tyrosine 3-hydroxylase  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 04-Dec-1986 #sequence\_revision 04-Dec-1986 #text\_change 03-Mar-2000  
 C:Accession: A00510; A44714; S03026; I58264  
 R:Grima, B.; Lamouroux, A.; Blanot, F.; Faucon Biguet, N.; Mallet, J.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 617-621, 1985  
 A:Title: Complete coding sequence of rat tyrosine hydroxylase mRNA.  
 A:Reference number: A00510; MUID:85113249  
 A:Accession: A00510  
 A:Molecule type: mRNA  
 A:Residues: 1-498 <GR>  
 A:Cross-references: GB:M10244; MID:g207408; PIDN:AAA42257.1; PID:g207409  
 R:Campbell, D.G.; Hardie, D.G.; Vuillet, P.R.  
 J. Biol. Chem. 261, 10489-10492, 1986  
 A:Title: Identification of four phosphorylation sites in the N-terminal region of tyrosine hydroxylase  
 A:Reference number: A44714; MUID:86278113  
 A:Accession: A44714  
 A:Status: preliminary  
 A:Molecule type: protein  
 A:Residues: 2-12;16-24;38-47;151-157 <CAM>  
 R:Bonnefoy, E.; Ferrara, P.; Rohrer, H.; Gros, F.; Thibault, J.  
 Eur. J. Biochem. 174, 685-690, 1988  
 A:Title: Role of the N-terminus of rat pheochromocytoma tyrosine hydroxylase in the regulation of its activity  
 A:Reference number: S03026; MUID:88271342  
 A:Accession: S03026  
 A:Molecule type: protein  
 A:Residues: 2-26 <BON>  
 R:Harrington, C.A.; Lewis, E.J.; Krzemien, D.; Chikaraishi, D.M.  
 Nucleic Acids Res. 15, 2363-2384, 1987  
 A:Title: Identification and cell type specificity of the tyrosine hydroxylase gene promoter  
 A:Reference number: I58264; MUID:87174758  
 A:Accession: I58264

A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-30 <RES>  
A:Cross-references: EMBL:X04914; NID:g57355; PIDN:CAA28584.1; PID:g57356  
C:Function:  
A:Description: catalyzes the 3'-hydroxylation of tyrosine to 3',4'-dihydroxyphenylalanine  
A:Pathway: catecholamine biosynthesis  
C:Superfamily: phenylalanine 4-monooxygenase  
F:8/Binding site: phosphate (Ser) (covalent) (by unidentified kinase) #status experiment  
F:19/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status exp  
F:40/153/Binding site: phosphate (Ser) (covalent) (by CAMP-dependent kinase) #status exp  
F:331,336,376/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.8%; Score 223; DB 1; Length 498;  
Best Local Similarity 26.7%; Pred. No. 2.5e-10;  
Matches 65; Conservative 39; Mismatches 103; Indels 36; Gaps 5;

QY 112 SSRFLWKS-----YCPRFLLDYLAFLGLSDFLDH-----QAVIKFELETHF 155  
DB 227 AEIATWKEVYVTLKGLYATHACHREHLEGFQLLERYCGYREDSTIPQLEDVSRFLKERTGF 286

QY 156 SYTPSGVAPHOYLSLDRPPIASVMTLDKDNFSLTDLHLGLHVPWLLHPSES 215  
DB 287 QLRPVAGLLSARDFLASLAFRVQCTQYIRHASSPMHSPEDPCCHELLGHVPMADRTFA 346

QY 216 EFFIMNGRLFTKVLKQVLPKSKKQRIOTLQSNLIAIVRCFWTFVSESLIENHEGRKAYG 275  
DB 347 QFSODIG-----LASLGVSDEIEKLSL-----VWFVTEFGLCKQNGELKAYG 390

QY 276 AVLSISPOLGHAFIDNVRLVPLELQDIIRLFNTSTPQETLFSIRHF---DELVELTSKLE 334  
DB 391 AGLSSYGEHLLSUSEPEVRAFDPTAAVQYQDQYQYVVFSEFSDAKOKLRSYASRIQ 446

QY 332 KLE 334  
DB 451 RIQ 453

RESULT 14  
I45983  
tyrosine 3-monooxygenase (EC 1.14.16.2) - bovine  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 16-Aug-1996 #sequence\_revision 16-Aug-1996 #text\_change 31-Mar-2000  
A:Accession: I45983  
R:Kawano, S.R.; Weisberg, E.P.; Stachowiak, M.K.; Turzai, L.M.; Gioio, A.E.; Kaplan, B.  
J. Neurosci. Res. 19, 440-449, 1988  
A:Title: Isolation and nucleotide sequence of a cDNA clone encoding bovine adrenal tyrosine 3-monooxygenase  
A:Reference number: I45983; MUID:88259287  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-491 <DN>  
A:Cross-references: GB:X36794; NID:g163750; PIDN:AAA30779.1; PID:g163751  
C:Superfamily: phenylalanine 4-monooxygenase  
C:Keywords: biotin; iron; metalloprotein; monooxygenase; oxidoreductase  
F:324,329,369/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.7%; Score 221.5; DB 2; Length 491;  
Best Local Similarity 27.1%; Pred. No. 3.2e-10;  
Matches 65; Conservative 42; Mismatches 102; Indels 31; Gaps 5;

QY 107 WYRLLSFRSLWKSYPFRFLDYLAFLGLSDFLDH-----QAVIKFELETHFSY 158  
DB 226 WKEVYVTLKGLYTHACR---EHLEAFELLERFCYREDRIPQLEDVSRFLKERTGFQLR 282

QY 159 PVSGVAPHOYLSLDRPPIASVMTLDKDNFSLTDLHLGLHVPWLLHPSESSEFF 218  
DB 283 PAAGLLSARDFLASLAFRVQCTQYIRHASSPMHSPEDPCCHELLGHVPMADRTFAQFS 342

QY 219 INMGRFLTKVLKQVLPKSKKQRIOTLQSNLIAIVRCFWTFVSESLIENHEGRKAYGAVL 278  
DB 343 QDIG-----LASLGVSDEIEKLSL-----LYWTFVTEFGLCKQNGELKAYGAGL 386

QY 279 ISSPOLGHAFIDNVRLVPLELQDIIRLFNTSTPQETLFSIRHF---DELVELTSKLE 334  
DB 387 LSSYGEHLLSUSEPEVRAFDPTAAVQYQDQYQYVVFSEFSDAKOKLRSYASRIQ 446

RESULT 15  
WHUY4  
tyrosine 3-monooxygenase (EC 1.14.16.2), splice form 4 - human  
N:Alternate names: tyrosine 3-hydroxylase  
N:Contains: tyrosine 3-monooxygenase, splice form 1; tyrosine 3-monooxygenase, splice  
C:Species: Homo sapiens (man)  
C:Date: 31-Mar-1989 #sequence\_revision 31-Mar-1989 #text\_change 16-Jun-2000  
A:Accession: A30002; A26825; A60201; JE0012; JE0013; JE0014; A27791; B27791; C27791;  
R:Nagatsu, T. GenBank, December 1987  
A:Reference number: A94509  
A:Accession: A30002  
A:Molecule type: mRNA  
A:Residues: 1-528 <NAG1>  
A:Cross-references: GB:M17589; NID:g339680; PIDN:AAA61179.1; PID:g339681  
R:Kaneda, N.; Kobayashi, K.; Ichinose, H.; Kishi, F.; Nakazawa, A.; Kurosawa, Y.; Fuji  
Biochem. Biophys. Res. Commun. 146, 971-975, 1987  
A:Title: Isolation of a novel cDNA clone for human tyrosine hydroxylase: alternative  
A:Reference number: A90136; MUID:87298614  
A:Accession: A26825  
A:Molecule type: mRNA  
A:Residues: 1-94 <NAG2>  
A:Cross-references: GB:M17589; NID:g339680; PIDN:AAA61179.1; PID:g339681  
R:Le Bourdellès, B.; Boularand, S.; Boni, C.; Horellou, P.; Dumas, S.; Grima, B.; Mal  
J. Neurochem. 50, 988-991, 1988  
A:Title: Analysis of the 5' region of the human tyrosine hydroxylase gene: combinator  
A:Reference number: A60201; MUID:88117543  
A:Accession: R60201  
A:Molecule type: mRNA  
A:Residues: 1-65 <LER>  
A:Cross-references: GB:M24790; NID:g556223; PIDN:AAA61174.1; PID:g556224  
R:Kobayashi, K.; Kaneda, N.; Ichinose, H.; Kishi, F.; Nakazawa, A.; Kurosawa, Y.; Fuji  
J. Biochem. 103, 907-912, 1988  
A:Title: Structure of the human tyrosine hydroxylase gene: alternative splicing from  
A:Reference number: JE0012; MUID:89008200  
A:Accession: JE0012  
A:Molecule type: DNA  
A:Residues: 1-30, 62-135 <KOB1>  
A:Cross-references: GB:D00269; NID:g220099; PIDN:BAA25094.1; PID:g2951764  
A:Experimental source: splice form 1  
A:Note: this splice form is produced by an alternative donor site within exon 1  
A:Accession: JE0013  
A:Molecule type: DNA  
A:Residues: 1-34, 62-135 <KOB2>  
A:Cross-references: GB:D00269; NID:g220099; PIDN:BAA25097.1; PID:g2951767  
A:Experimental source: splice form 2  
A:Accession: JE0014  
A:Molecule type: DNA  
A:Residues: 1-30, 35-135 <KOB3>  
A:Cross-references: GB:D00269; NID:g220099; PIDN:BAA25095.1; PID:g2951765  
A:Experimental source: splice form 3  
A:Note: this splice form is produced by an alternative donor site within exon 1  
R:Grima, B.; Lamouroux, A.; Boni, C.; Julien, J.F.; Javoy-Agid, F.; Mallet, J.  
Nature 326, 707-711, 1987  
A:Title: A single human gene encoding multiple tyrosine hydroxylases with different p  
A:Reference number: A93393; MUID:87173064  
A:Accession: A27791  
A:Molecule type: mRNA  
A:Residues: 1-30, 62-528 <GR11>  
A:Cross-references: GB:X05290; NID:g32501; PIDN:CAA28908.1; PID:g32502  
A:Experimental source: splice form 1  
A:Note: this splice form is produced by an alternative donor site within exon 1  
A:Accession: B27791  
A:Molecule type: mRNA

A:Residues: 1-34,62-528 <GRI2>  
A:Cross-references: GB:X05290; NID:g32501  
A:Experimental source: splice form 2  
A:Accession: C27791  
A:Molecule type: mRNA  
A:Residues: 30,35-528 <GRI3>  
A:Cross-references: GB:X05290; NID:g32501  
A:Experimental source: splice form 3  
A:Note: this isoform is produced by use of an alternative donor site within exon 1  
R:Ichinose, H.; Ohye, T.; Fujita, K.; Yoshida, M.; Ueda, S.; Nagatsu, T.  
Biochem. Biophys. Res. Commun. 195, 158-165, 1993  
A:Title: Increased heterogeneity of tyrosine hydroxylase in humans.  
A:Reference number: PNO575; MUID:93371398  
A:Accession: PNO575  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 19-30 <ICH1>  
A:Accession: PNO582  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 35-61 <ICH2>  
A:Accession: PNO588  
A:Status: translation not shown  
A:Molecule type: DNA  
A:Residues: 62-106 <ICH3>  
R:O'Malley, K.L.; Anhalt, M.J.; Martin, B.M.; Kelson, J.R.; Winfield, S.L.; Ginns, E.I.  
Biochemistry 26, 2910-2914, 1987  
A:Title: Isolation and characterization of the human tyrosine hydroxylase gene: identification of a full-length cDNA clone encoding human tyrosine hydroxylase type  
A:Reference number: 138340; MUID:87316931  
A:Accession: 138340  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-30,35-528 <KOB4>  
A:Cross-references: EMBL:Y00414; NID:g37126; PIDN:CAA68472.1; PID:g37127  
R:Ginns, E.I.; Rehavi, M.; Martin, B.M.; Weller, M.; O'Malley, K.L.; Lamarca, M.E.; McAL  
J. Biol. Chem. 263, 7406-7410, 1988  
A:Title: Expression of human tyrosine hydroxylase cDNA in invertebrate cells using a baculovirus expression system.  
A:Reference number: 155282; MUID:88213428  
A:Accession: 155282  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-30,62-64 <GIN1>  
A:Cross-references: GB:M20911; NID:g339636; PIDN:AAA61167.1; PID:g339637  
A:Accession: 170056  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-34,62-64 <GIN2>  
A:Cross-references: GB:M20912; NID:g339642; PIDN:AAA61168.1; PID:g339643  
A:Comment: The expression of the four distinct proteins produced by alternate splicing of the tyrosine hydroxylase gene.  
C:Genetics:  
A:Gene: GB:TH  
A:Cross-references: GDB:119612; OMIM:191290  
A:Map position: 1p15.5-1p15.5  
A:Introns: 34/3; 61/3; 135/3  
A:Note: the list of introns is incomplete  
C:Function:  
A:Description: catalyzes the 3'-hydroxylation of tyrosine to 3',4'-dihydroxyphenylalanine  
A:Pathway: catecholamine biosynthesis  
A:Note: this is the rate-limiting step in catecholamine biosynthesis  
C:Superfamily: phenylalanine 4-monooxygenase  
C:Keywords: alternative splicing; biotin; catecholamine biosynthesis; iron; metalloproteinase  
F:1-528/Product: tyrosine 3-monooxygenase, splice form 4 #status predicted <MAT4>  
F:1-34,62-528/Product: tyrosine 3-monooxygenase, splice form 2 #status predicted <MAT2>  
F:1-30,35-528/Product: tyrosine 3-monooxygenase, splice form 3 #status predicted <MAT3>

F:1-30,62-528/Product: tyrosine 3-monooxygenase, splice form 1 #status predicted <MAT1>  
F:8/Binding site: phosphate (Thr) (covalent) (by unidentified kinase) #status predicted  
F:19/Binding site: phosphate (Ser) (covalent) (by calmodulin-dependent kinase) #status predicted  
F:71,183/Binding site: phosphate (Ser) (covalent) (by cAMP-dependent kinase) #status predicted  
F:361,366,406/Binding site: iron (His, His, Glu) #status predicted

Query Match 11.6% Score 220; DB 1; Length 528;  
Best Local Similarity 26.7%; Pred. No. 4.7e-10;  
Matches 65; Conservative 38; Mismatches 104; Indels 36; Gaps 5;

QY 112 SRSFSLWKS-----YCPREFDYLEAFGLLSDFLDH-----QNAVTFPELETHF 155  
DB 257 AEELATWKEVYITLKLATACGHELEAFALLERTSGYREDNIPOLVDVSRFLKERTGF 316  
QY 156 SYTPVSGFVAPHQYLSLQDRYFPPIASVMFTLDKDNFSLTPDLIHLGLHVPMLHPSPS 215  
DB 317 QLRPVAGLLSARDELASLAFRVFOCTQYIRHASSPMHSPPEPCCHHELLGHVPMADRTFA 376  
QY 216 EFFINMGRILTKVIEKVOALPSKKORIOTLOSNIATVRCFWFTVESGLIENHEGRKAY 275  
DB 377 QFSQDIG-----LASLGASDEEIEKLTLS-----WTFVEFLCKQNGEVKAY 420  
QY 276 AVLISSPOELGHAFIDNVRVLPFLDQILRIPFNTPSTPQETLFSIRHF----DELVELTS 331  
DB 421 AGLSSYVGLLHCLSEPEIRAFDPEAAAVQPYQDOTYQSVFVSEFSFSDAKDKLSRYAS 480  
QY 332 KLE 334  
DB 481 RIQ 483

Search completed: June 29, 2001, 13:54:16  
Job time: 805 sec

Mon Jul 2 08:46:44 2001

us-09-438-185-1047.rpr

Page 8